Logo Detection and Recognition from Scanned Documents

Sushama Khanvilkar[1], Vinay Billa[2], Saikirankumar Dasari[3], Saqib Naikwadi[4] , Aakash Tiwari[5]

Computer Department, Xavier Institute of Engineering, Mumbai, India

[1]sushma.k@xavierengg.com

[2]billavinay2012@gmail.com

[3]saikiran.dasari2011@gmail.com

[4]sinxie007@gmail.com

[5]aakash2997@gmail.com

***Abstract*--We present a randomized and heuristic algorithm based on computer vision principles to detect and recognize logos in documents such as invoices and purchase orders. It is intended for use in enhancing document retrieval (DR) and named entity recognition (NER) tasks. Our algorithm operates in two phases: first, it detects the area in a page of a scanned document where a logo appears, and marks this area by using a bounding box; then, it recognizes the organizational entity associated with the logo by finding the best match for it against a pre-existing database of known logos. We extract distinctive image features from both logo and document using the Scale Invariant Feature Transform (SIFT) algorithm and use the Fast Library for Approximate Nearest Neighbors (FLANN) for matching against the database, with a number of custom heuristics to reduce detection and recognition error. The technique is robust to image transformations such as rotation and scaling and to noise from blurriness and scanner artifacts.**

***Keywords*— FLANN, SIFT, Document Retrieval(DR), Named entity recognition(NER)**

I. INTRODUCTION

Public organizations, institutes, companies and private sectors are generally interested in implementing digital mail rooms to improve the efficiency of paper-intensive workflows and to reduce the burden of manual processing of different administrative documents including incoming mails, faxes, forms, invoices, reports, employee records, health record, etc. Logo can be considered as an important and popular salient entities presented in administrative documents. The manual identification/verification of logos is not an easy task, as the documents in-flow in organizations is growing rapidly.

Indeed, accurate detection and recognition of logo in document images provide us with a more reliable and appropriate system. However, logo detection/recognition is a challenging task, as logos are generally composed of quite complex symbols, graphical and textual components.

The main aim of this project is to present the efficient and robust framework for detection as well as recognition of logo images.

A. Aim and Objective :

The aim is to provide efficient and robust framework for detection as well as recognition of logo images.

The objectives are:

1) To present the literature review over different approaches presented over logo.

2) To present the analysis of different methods according to their detection accuracy and performances.

3) To present and discuss the proposed methods for logo detection and recognition.

4) To present the practical analysis of proposed work and its evaluation against the existing methods.

II. RELATEDWORK

The literature on Logo Detection and Recognition is extremely extensive. Financial transactions for exchange of goods or services between parties usually generate data in the form of physical documents such as invoices and purchase orders. Such documents are used to record, verify, and scrutinize transactions; they may either have a rigidly structured layout like a form, or be more unstructured in nature. Use cases such as domestic and international trade finance transactions typically involve additional third parties like banks and financial institutions as intermediaries who facilitate the exchange and act as trusted parties. While such parties earn significant revenue from such trade activities, they need to avoid invalid transactions, which may have legal sanctions and lead to significant losses.

To validate transactions, such intermediaries currently employ and invest in manual verification, which is painfully slow and inefficient. It is, therefore, of great interest to them to be able to leverage automated DR and NER. Techniques from such unstructured documents for use by authorized personnels or officials to validate and act on transactions. The work described here is a component of a larger platform-based solution to address this gap: to help visualize information, to improve decision making, and further minimize the processing time and increase productivity for this task.

Logos are an important element of documents, and detecting and recognizing them significantly augments the richness of the information being presented to the human operator. However, logos are not text, and are therefore not captured by traditional Optical Character Recognition (OCR) techniques. Specialized image-based techniques are therefore needed.

Various public cloud-based web services (such as Google Cloud Vision and ClarifAI [1]) support logo detection by exposing an API over REST-style frameworks. This saves considerable investment on model building activities (gathering training data, building neural network architectures, and fitting them to available data) at the cost of requiring documents to be uploaded to the public cloud. This is not considered an acceptable tradeoff for sensitive documents relating to financial transactions, where a data breach or information leak can result in significant losses or liability actions. This consideration ruled out for us the possibility of using pre-built models and transfer learning techniques, and led us to develop an on-premise robust algorithmic solution.

The problem of using logo information for document image processing basically involves two main tasks:

1) Finding boundary of a logo/ on a document image irrespective of its class.

2) Indexing/matching the detected logo/ candidate region to a database for classifying or for concluding that the region is not of interest. The former is referred to as logo/ detection/spotting, while the latter is called logo/ recognition.

III. DESIGN AND IMPLEMENTATION

The project aims at developing a Logo Detection and Recognition system from Scanned Documents which can be used in various organizations, institutes, etc. The technology is rapid efficient and robust.

For computing key-points, we use the OpenCV implementation of SIFT. For key-point matching we use the OpenCV implementation of the Fast Library for Approximate Nearest Neighbour Search (FLANN) [11], which provides a set of algorithms to do fast nearest-neighbour search for high-dimensional features.

A. Feature Detection and Extraction:

We use principles of object detection and tracking to match common features between images. The unique patterns in an image or the points of interest which describe the image are known as key-points [8]. Detection of this key-points will be done using an algorithm which are invariant to scale and rotational transformation. We have compared SURF [6], SIFT, BRISK [9] and KAZE [5] algorithms and we found SIFT algorithm is best suited for all the above specifications. SIFT (Scale Invariant Feature Transform): SIFT uses following four-step process: Scale-space extrema are detected based on a Difference-of-Gaussians (DoG) filter, which approximates a Laplacian-of-Gaussians (LoG) operator. Key-points are localized by eliminating and rening low contrast key-points and edge key-points. The orientation of the key-point is assigned by calculating the scale, gradient magnitude, and direction in the region of the neighbourhood. The descriptor is calculated using 16 sub-blocks around the key-point neighbourhood, each divided into 4x4 size. For each sub-block, an eight-bin orientation histogram is created. This descriptor is represented as a 128-dimensional vector.

B. Feature Matching:

Once we have extracted features for both logos and pages, we need to find a “best match " between the features of a page and those of logos in the logo bank. We here need a brute force matching using hierarchical fast nearest-neighbour search. Lowe [10] uses a nearest-neighbor technique to compute the top two matches, and then uses the ratio of the closest distance to the second-closest distance to reject matches that exceed a threshold. This matching technique is low cost, fast, and effective when compared to the brute-force approach.

C. Homographies and Outlier Filtering:

After determining the matching features or key-points between the page and a logo, we locate the logo on the page by trying to fit a transformation with respect to the key-points. The transformation is denied by a homography matrix, a 33 matrix H giving the perspective transformation of the matched area with respect to the logo image. In case no homography matrix is found for a set of key-points, we conclude that no logo is present on the page. Although Lowe's ratio test removes most incorrect matches, there can still be outliers in the match. The Random Sample Consensus (RANSAC) estimator is a probabilistic method of removing outliers from matched features that can approximate the homography matrix H even with very few matched points. Finally, we compute the perspective transformation using the matched points and the homography matrix H. This returns a set of points which represents the pixel coordinates on the page of the bounding box of the logo.

D. *List of Heuristic Checks:*

The data challenges discussed earlier introduce errors which are not solved by techniques like homography estimation. While a homography matrix H can be computed between the key-points on the logo and the document, a meaningful match should conform to certain physical realities. We use the characteristics of a good match as our basis for eliminating false positives and increasing solution accuracy, and now list a comprehensive set of heuristic checks for this purpose. All of these checks are Boolean in nature.

1) *Matched Area Test:*

We do not expect the matched logo to occupy the majority of the space on the target page. The ratio of these areas must be below a threshold TM.

2) *Homography Matrix Validation Test:*

The homography matrix which defines a transformation can be computed mathematically but may not be valid as per traditional image transformation rules. One such way of validating a homography matrix is to compute its determinant. If the determinant value approaches zero, then the homography matrix is singular, which means we are seeing the plane object at 90 degrees, which is impossible. Likewise, the ratio of the matched area on the page to the matched logo area (given by the square of the determinant) should be a finite number and not exceed a threshold. Intuitively, the transformation can expand or shrink the logo in document, but not by a very large factor. Additionally, if the determinant of the homography matrix is negative, then the homography is not conserving the orientation, except for mirror images, which is outside the scope of this paper. SIFT and similar algorithms are not known to be mirror-image invariant.

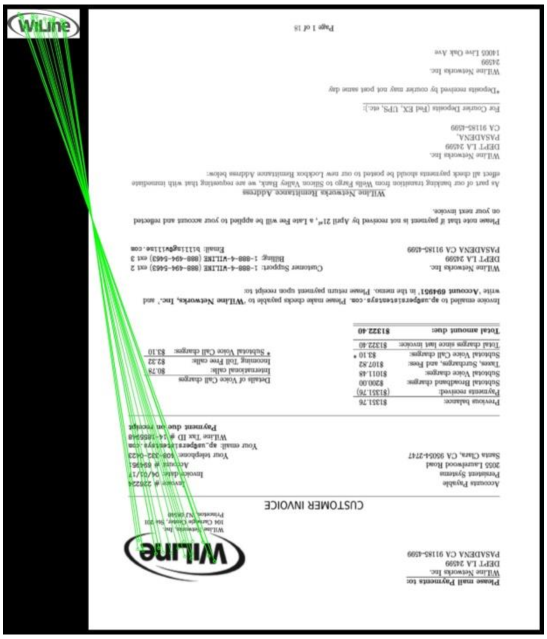

Fig. 1 Example match result on the same document scanned up-side down

3) Aspect Ratio Test:

The aspect ratio should ideally be preserved by the homography transformation, for a good match. The computed ratio of the aspect-ratios of the logo and its image on the page should not exceed a threshold TA. We allow the tolerance to take into consideration the changes caused by perspective transformation and introduced noise.

4) Border Pixel Match Test:

The key-point match is struck down if the matching key-points are at or within TB pixels of the border of the page. This is equivalent to shrinking the match area by adding margins.

E. Putting It Together:

We use open-source tools like ImageMagick [2] and associated Python packages for preprocessing document images and logos. All development is done in Python, using the Anaconda environment with Jupyter Notebooks [3]. The Open Source Computer Vision library (OpenCV) [7] built with Python is used for data ingestion, feature detection, and various image processing tasks. We pre-compute and store the key-points of all logos in the logo bank, along with metadata such as the logo's height and width. Logos are resized for maintaining correct aspect ratios. An arbitrary logo downloaded from the internet or captured from documents may contain uninformative whitespaces, so that even a rectangular shaped logo would appear square. The logos are cropped to maintain the correct aspect ratio as needed for the match. If needed, logos are compressed as well for optimizing size. We then convert them or standardize them to gray scale formats from color (RGB) formats. We use the Mogrify program of the ImageMagick suite to resize and strip useless information and plane-interlace the color channel values.

Financial documents scanned and uploaded from various endpoint scanners, such as at supplier locations, vendor offices, and local branches of financial institutions are collected using existing Enterprise Resource Planning (ERP) systems. We first preprocess these PDF files using ImageMagick tools to convert them to an image format and to optimize for size if needed. For some cases we also apply a very small Gausian blur (radius = 0.05), as this enhances the detection of logo features, while making it less likely to detect them in textual content area. The page images are then converted into gray scale format from color (RGB) formats.

For computing key-points, we use the OpenCV implementation of SIFT. For key-point matching, we use the OpenCV implementation of the Fast Library for Approximate Nearest Neighbor Search (FLANN) [17], which provides a set of algorithms to do fast nearest-neighbor search for high-dimensional features. We run our heuristic checks for each logo A clearing Lowe's ratio test for a given page p, requiring unanimous consensus from all of the heuristics in order for to A to proceed further. Should multiple logos still survive, we sort them by decreasing order of their key-point match confidence, using the value of abs(det(H)) to break any remaining ties.

IV. RESULTS

We now discuss the performance of our solution on a set of multi-page PDF documents like invoices and purchase orders containing distinct logos. Of these, logos were graphical, textual, and mixed. No page contained more than a single logo. We selected pages from the documents, each containing either zero or one logo from our logo bank. The rise in accuracies shows the critical role of the heuristics in the solution. As shown in Figure 1, the matches are rotational invariant and independent of the degree of rotation.

We emphasize that these results are necessarily specific to the data bank and the set of pages, and extrapolation to other scenarios should be performed only with great caution.

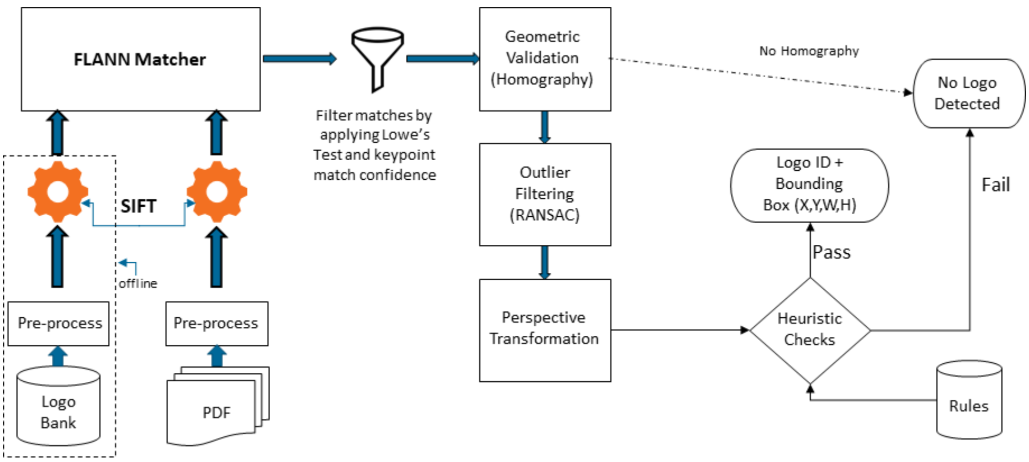

Fig. 2 Solution process flow diagram.

The model is fed with the input and the logo is retrieved. Fig. 3 shows the output for a given document.

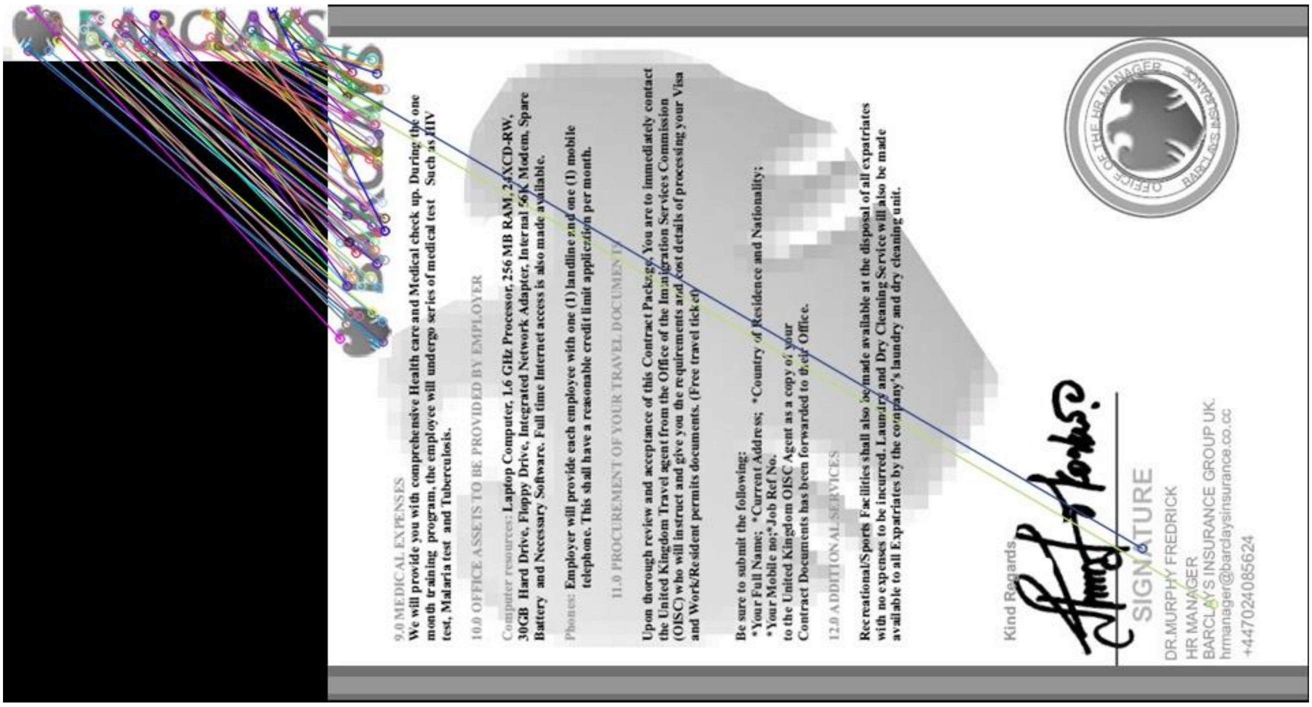

Fig. 3 Logo detection

V. CONCLUSION

We have eliminated various feature detection techniques such as SURF (Speeded Up Robust Features) and BRIEF (Binary Robust Independent Elementary Features) as we found SIFT to be well suited for both scale and rotationally-invariant changes of logos from the scanned documents. We have studied and evaluated various feature detection and description methods coupled with corresponding matching algorithms for extracting and recognizing logos in financial documents such as invoices, bills of lading, and purchase orders. The combination of SIFT for key-point detection and extraction and FLANN for key-point matching gave us the best results. The use of additional heuristics vastly improves algorithm accuracy and minimizes false positives. Our solution is currently being developed to augment various use cases such as DR and NER.

VI. FUTURE WORK

A. Our various heuristics (especially their threshold values) are sensitive to various characteristics of the logos and the documents. Rather than setting their values by trial-and-error as we have currently done, they can and should be studied as data-driven parameter search problems. Hyperparameter search techniques used in machine learning should be used to find the best values for the thresholds T_M, T_N , T_L, T_U , T_A, and T_B , assuming availability of sufficient training examples.

B. The extension to multiple logos on a single page is not conceptually difficult in the case where we have multiple matches in which we have good confidence and whose homography matrices place the logos in non- overlapping areas of the page. The efficacy of this approach needs to be validated empirically.

C. The current solution is not closed-loop, in the following sense: if a page contains a previously unseen logo (i.e., one currently not existing in the logo bank), this can result either in a false positive or in no match, but there is no way to distinguish these outcomes and use it to provide (possibly human) feedback that would allow the system to bootstrap itself. The problem of incremental corpus enlargement is not unique to this usecase.

D. We can consider using information from color channels, where available, to improve accuracy. The additional information could be useful for creating more discriminating descriptors, and could further mitigate the problem such as incorrect borders and occlusions by seals.

E. Seal detection, which has to address a very different set of problems (see [4]), can provide additional information on entities such as government organizations, ports, and processing entities, which would all be useful in this application domain.

F. Given its high level of accuracy, our solution is well- suited for use as a test set generation and annotation scheme for various deep learning-based algorithms for Logo Detection problems.

ACKNOWLEDGMENT

We express sincerest gratitude to our project guide, Head of Department Sushama Khanvilkar, who has been a source of inspiration and has made it possible for us to initiate and carry on with this project.

REFERENCES

[1] ClarifAI Logo Model. https://www.clarifai.com/models/.Accessed on 2018-08-18.
[2] ImageMagick. https://github.com/ImageMagick/ImageMagick. Ac- cessed on 2018-08-18.
[3] Jupyter Notebook. http://jupyter.org/. Accessed on 2018-08-18.
[4] Alaei, A., Roy, P., and Pal, U. Logo and seal based administrative document image retrieval: A survey. Computer Science Review Vol. 22 (09 2016), 47–63.
[5] Alcantarilla, P. F., Bartoli, A., and Davison,A.J. Kaze features. In Proceedings of the 12th European Conference on Computer Vision - Volume Part VI (Berlin, Heidelberg, 2012), ECCV'12, Springer-Verlag, pp. 214–227.
[6] Bay, H., Ess, A., Tuytelaars, T., and Van Gool,L. Speeded-up robust features (SURF). Comput. Vis. Image Underst. 110, 3 (June 2008), 346–359.
[7] Bradski, G. The opencv library. Dr. Dobb's Journal of Software Tools (2000).
[8] Hassaballah, M., Ali, A., and Alshazly, H. Image Features Detection, Description and Matching, vol. 630. Springer, Cham, 02 2016.
[9] Leutenegger, S., Chli, M., and Siegwart, R. Y. Brisk: Binary robust invariant scalable keypoints. In Proceedings of the 2011 International Conference on Computer Vision (Washington, DC, USA, 2011), ICCV '11, IEEE Computer Society, pp. 2548–2555.
[10] Lowe, D. G. Distinctive image features from scale invariant keypoints. Int. J. Compute. Vision 60, 2 (Nov. 2004), 91–110.
[11] Muja, M., and Lowe, D. G. Fast approximate nearest neighbors with automatic algorithm configuration. In International Conference on Computer Vision Theory and Application VISSAPP'09) (2009), INSTICC Press, pp. 331–340.