EURASIAN JOURNAL OF MATHEMATICAL AND COMPUTER APPLICATIONS
ISSN 2306–6172
Volume 13, Issue 4 (2025) 85– 92

OPTIMIZED TEXT RETRIEVAL FRAMEWORK FOR LOW-QUALITY PRINTED DOCUMENTS

Habeeb I.Q. [1], **Al-Zaydi Z.Q.**, **Abdulkhudhur H.N.**

Abstract Resolution is always a key factor for all types of image processing applications. One common application is text retrieval from images. The assumption that all input images are of good quality may not be met in some cases. Low-quality images can be taken from a low-resolution clip, old printed documents, or downloaded from one of the millions of documents available online. Traditional image preprocessing techniques that use a super-resolution algorithm are mostly designed to enhance natural scenes and face detection. One of the biggest challenges they face is the possibility of losing certain textual details in the noise reduction process. In this study, an optimized framework is specifically designed for text retrieval from low-resolution documents. It can recover missing textual content, which is often eliminated by existing denoising methods. The proposed framework extracts useful features from a small number of outputs for the same low-resolution image. The outputs are generated using four interpolation algorithms. The proposed framework is evaluated on test images scanned at different low resolutions. The evaluation proves the superiority of our proposed framework over other text retrieval methods, which is expressed by a better accuracy compared to the best existing method by 11%.

Keywords: text extraction, low-resolution image, interpolation algorithm, old printed document, historical document recognition.

AMS Mathematics Subject Classification: 68W01, 68W32, 68T50.

DOI: 10.32523/2306-6172-2025-13-4-85-92

1 Introduction

Digital media and information technology have recently evolved, greatly increasing the amount of information that can be captured and shared, including photos and videos [1]. Retrieving such text data from a huge volume of images and video requires an efficient approach [2]. For this reason, in recent years, computer vision and multimedia applications have relied heavily on text retrieval systems [3]. Text retrieval is a broad research topic involving many fields [4], such as information retrieval [5], computer vision [6], machine learning [7], and human-computer interaction [8]. With the millions of images out there today, it is almost impossible to manually annotate each individual image [4]. Text retrieval techniques have been developed to overcome these limitations. These techniques provide faster text-based extraction within images. However, poor-quality or low-resolution images cause resolution issues that severely restrict automatic or text usage [9]. Old printed documents, online downloads, or photos taken with low-resolution devices such as webcams and cell phones can produce low-resolution images.

High resolution makes the image sharper, clearer, and more detailed [10]. In contrast, low-resolution images lack enough pixels to fully capture the nuances of text, which is why they are very blurry [11]. This makes it difficult to distinguish between the shapes of these letters because many of them appear to be overcoming each other. Text segmentation and

[1]Corresponding Author.

feature extraction methods may become more difficult with such low-quality images [12]. For example, during the text segmentation process into its letters, noise may appear in many character blocks. Sometimes, a letter may appear in one block and the rest in a different block. In other cases, two or more influential letters appear in one block. Using super-resolution algorithms to pre-process images is a traditional method that faces major difficulties [13]. Typically, they are made to detect faces and natural scenes. Moreover. After noise reduction, texture details within the image may always be lost [14]. Furthermore, unless the original scan resolution is high, enhancing the image resolution after scanning will not add additional information to the image [15]. Fig. 1 shows the effect of image resolution on text retrieval accuracy.

In Fig. 1, text retrieval values for different image resolutions were recorded using the Google Tesseract engine [16]. The values prove that the accuracy of text retrieval often decreases due to the poor quality of input document images. This means that before sending document images to the text retrieval engine, further work must be done to increase their quality. Interpolation is a frequently used method to improve image resolution [17]. It is a straightforward technique for improving image dimensions by simply adding new pixels or data points to the low-resolution image. Commonly used image interpolation algorithms include B-Spline, Bi-cubic, Bi-linear, and Nearest Neighbor [18]. Due to the effect of interpolation, the high-frequency components of the image, such as edges, are the primary loss area when the image resolution increases due to interpolation. Edges are critical for detecting objects within an image [19]. Therefore, it is important to preserve all edges of the image, which is essential to improve image quality.

Applying the interpolation methods produces different results because each method rescales the image in a different way [20]. Therefore, this study proves that it is possible to take advantage of outcome features resulting from different interpolation methods to obtain better results. Furthermore, this research has developed a practical approach that can be applied to any available low-resolution image to enhance and restore its quality so that text retrieval systems can easily recognize it. Details of the proposed framework are provided in section three.

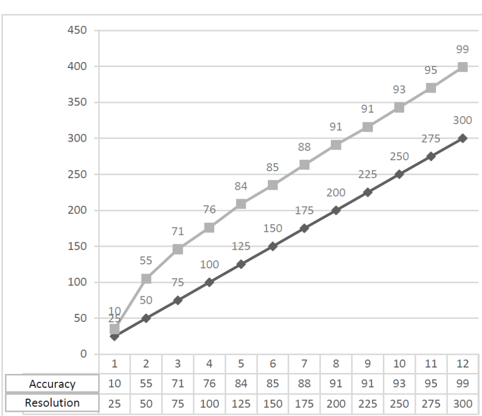

Figure 1: Comparing text retrieval accuracy to different image resolution.

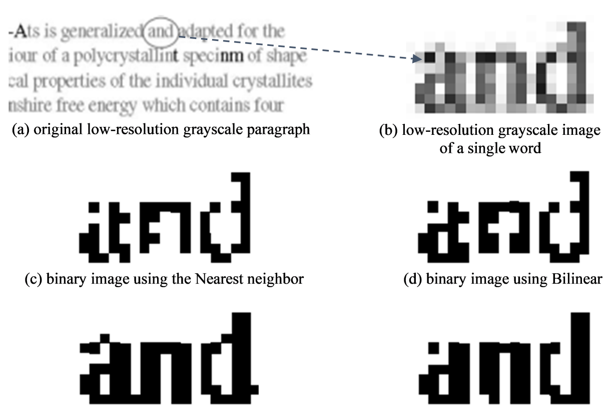

Figure 2: Binary images using four types of interpolation algorithms.

2 Related work

This section examines the attempts made by other researchers in trying to address the research problem. The focus is on the methods that have created solutions, how they work, and their strengths and weaknesses. In [21], information was extracted from low-quality medical images using a Bi-cubic interpolation technique. This was achieved by measuring, evaluating, and comparing the results with Nearest-neighbor and Bi-linear algorithms using three parameters: root mean square, mean square error, and peak signal-to-noise ratio. The experimental results indicate that the Bi-cubic approach performs better in terms of accuracy than the Bi-linear and Nearest-neighbor algorithms. Furthermore, the degree of similarity to the input image grows with the size of the image resolution, but at the cost of increased computational complexity.

Some researchers, such as [22], claimed that there are many problems associated with isolated character recognition using text retrieval systems. One such problem is that the accuracy of isolated letter recognition decreases dramatically as font sizes decrease. After using different super-resolution techniques on images of smaller, isolated letters, the researchers compared recognition accuracy and examined how recognition accuracy changed with image size. They used highly accurate algorithms based on deep learning models, such as SRCNN and DCSCN, to recognize isolated Korean characters with small fonts. The Korean-language scanned images make up the testing dataset. According to the test results, the most efficient method for reducing the error rate is to use DCSCN super-precision. The outcomes of the experiments indicate that including super-resolution techniques in the preprocessing module can enhance the precision of text recognition for smaller Korean fonts.

Recently, text recognition in natural scenes has been investigated. For example, [23] showed the importance of text retrieval in a wide range of vision applications. Hence. They summarize the major issues associated with scene text recognition. Including the effect of low-resolution images on recognition accuracy. Handwriting recognition based on training data was seen as a challenging task in [24]. The authors claimed that as the amount of handwritten data increases, deep learning techniques should be the main focus in the field of handwriting recognition. They presented that convolutional neural networks are the best option for dealing with handwriting recognition problems due to their ability to determine the structure of handwritten characters in a way that facilitates automatic extraction of useful features. They optimize the structure of the convolutional neural network by adjusting the kernel size, learning parameters, and number of layers. They ran a lot of tests and on a dataset of handwritten texts, they were 99% accurate in recognizing them.

In [25], irregular text is considered largely difficult to recognize due to its varied outlines and distorted forms. Therefore, a multi-object detection system is designed for text recognition. Its design consists of a detection and correction network. The system detects and extracts irregular text from scene images. The design reduces the complexity of the recognition process and allows irregular text to be extracted more easily. However, the design requires training with supervision of images and their text labels. Multi-object detection system focuses on isolated letters and shows the result sequentially. Besides, the system can recognize regular and irregular text-embedded scene images. Many experiments were conducted on different test databases, which showed that the proposed system achieves high accuracy.

Authors in [26] demonstrated that the task of low-resolution face recognition remains challenging, especially when faces are taken under non-perfect environments, which is commonly prevalent in vision-based applications. Hence, they presented a face recognition framework for low-resolution images. The framework can handle non-ideal situations, such as non-uniform lighting and non-frontal face positions. They mentioned three contributions, (1) conducting tests to evaluate low-resolution techniques for face recognition, (2) studying face restoration on different public datasets, including large-scale datasets, providing the fundamental outcome for existing deep learning neural approaches, and (3) exploring face detection using state-of-the-art supervised learning approaches. Recent investigative work mentioned above demonstrates a variety of approaches and initiatives used to enhance text recognition of low-resolution images. However, as the current section shows, the bulk of them overlooked the proposed framework of this study.

3 Proposed Framework

The idea of the proposed framework is based on the fact that a text retrieval system is a software procedure and always produces the same output under the same input image file and constraints. Therefore, it is possible to change the pixel values of an image slightly to get a different digital file and different output. This can be achieved by extracting useful features from a small number of outputs using different interpolation algorithms for the same low-resolution image. Fig. 2 shows different binary images using four types of interpolation algorithms for the same gray-scale image.

In Fig. 2, the low-resolution image of a paragraph is shown in (a) while the single word "and" is isolated in (b). The image in (b) has been re-scaled using four interpolation algorithms: Nearest neighbor, Bi-linear, Bi-cubic, and B-Spline. After that, four binary images of the outputs of the interpolation algorithms are represented in (c), (d), (e), and (f) using a threshold value of 130. These four binary images are slightly different and may not be clear to the human eye. But it certainly produces different features and thus may lead to another segmentation and classification of letters. For example, the image in (c) is not the same as the image in (e) due to some letters appear to be touching while others are separate. This is because the gray-scale values of the pixels in both do not exactly match resulting in different features. Generating N number of versions of the input image using four different interpolation algorithms is the core component of this approach. Then, the best features of these versions can be selected for the final text retrieval system output. The proposed framework and its stages are shown in Fig. 3.

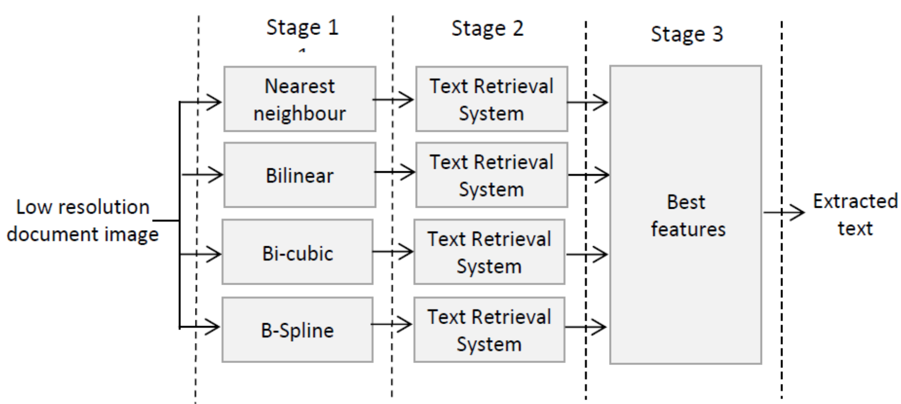

Figure 3: Proposed framework.

Fig. 3 shows that the proposed framework consists of three main stages. In stage 1, the low-resolution document image is passed to four interpolation algorithms in parallel to re-scale its resolution in different ways. These four interpolation algorithms are Nearest Neighbor, Bi-linear, Bi-cubic, and B-Spline. According to researchers [27], they are the best image-resizing algorithms to use for preprocessing before passing the low-resolution document to the text retrieval system. Details of these methods and how they work are described in the following subsections.

3.1 Nearest Neighbors Method

Making each pixel in the new image equal to its closest pixel in the original image is one of the easiest techniques for up-scaling an image. Although this approach is computationally fast, it can make the image look fuzzy, especially if the values of a large number of pixels are changed. Images that contain duplicate pixels may lose some of their information. When you want objects within the image to have sharp edges, this method is ideal. For example, if the values of point P1 and point P2 are 6 and 12 respectively. Then point P3 becomes equal to 6 if it is closer to point P1 [28].

3.2 Bi-linear Method

Bi-linear interpolation is a more complex method for re-sampling images that determines pixel values by taking into account a weighted mean of the surrounding points in the source image. In contrast to the nearest neighbor, which only takes into account the nearest pixel, it looks at neighboring pixels from both horizontal and vertical directions. It uses linear equations to calculate the weighted mean and blends the color values of nearby pixels according to how far they are from the target point. This results in a higher-quality, scaled image by creating a smoother transition between pixels [29].

3.3 Bi-cubic Method

When speed is not a concern, bi-cubic interpolation is frequently preferred in image processing over bi-linear or nearest-neighbor methods in image re-scaling. Bi-cubic interpolation considers 16 pixels (4x4), but bi-linear interpolation only examines 4 pixels (2x2). The closest pixels receive more attention in calculations and are an ideal combination of output quality and processing time. For this reason, this method is standard in many image processing programs, such as Adobe Photoshop. All objects within an image will appear more continuous, even when they're not, and it probably generates more accurate pixel values overall [27].

3.4 B-Spline

This method is a polynomial function that uses a set of flexible groups controlled by a set of values called control pixels, which create smooth isolated characters. These flexible groups are used to produce clear images and shapes using a specific set of pixels. The places where curves meet are known as nodes. B-spline is primarily designed to represent a curve due to its attractive characteristics. For example, boundaries, continuity, spatial singularity, and preservation of structure. Hence, it has been adopted for use in object segmentation and classification and thus has been widely used in image processing [30].

Returning to Fig. 3, in Stage 2 of the proposed framework, each text retrieval engine will receive a single re-scaled image from one of the four interpolation algorithms used in this research. These versions of the re-scaled image are similar but not equal. Finally, the words inside the re-scaled image are extracted and stored as a fixed array of words. Then, in Step 3, the words of each array are organized in parallel with their equivalents in other text retrieval outputs using the Smith-Waterman technique. The best features (words) of each column will be selected to generate the final text retrieval output by combining the candidate words using a space as a separator. The best words were elected and marked as valid by examining each resulting word if it was defined in a dictionary. Otherwise, if no word is defined in the dictionary, the word with the largest number of letters in common is chosen.

4 Empirical Results

The proposed framework is evaluated against four popular interpolation algorithms, including Nearest Neighbor, Bi-linear, Bi-Cubic, and B-Spline. An interface is programmed and pro-

Table 1: Proposed framework evaluation.

Input Images	Nearest neighbor	Bilinear	Bi-cubic	B-Spline	Proposed Framework
Images with 50 dpi	46%	54%	56%	61%	79%
Images with 75 dpi	65%	73%	74%	77%	88%
Images with 100 dpi	74%	77%	79%	81%	93%
Images with 150 dpi	78%	81%	82%	84%	95%

duced using Visual studio.NET that includes procedures to execute the proposed framework and interpolation algorithms. The Tesseract engine is also used as a text retrieval system to extract text from images. Tesseract is a freely available software tool from Google that can be integrated into Visual Studio. NET. Experiments with a total word count of 38958 have been conducted on many English documents. The texts in these documents serve as reference text and are taken directly from the CNN News Network website without any layout or pictures. The text is first printed on paper, and then the hard-copy is scanned at 50 dpi, 75 dpi, 100 dpi, and 150 dpi with an 8-bit gray-scale to create low-resolution test images from the reference. The comparison result of the accuracy measure for different interpolation algorithms is shown in Tab. 1.

Tab. 1 shows the superiority of the proposed framework over existing interpolation algorithms in the accuracy measure. Accuracy is equal to dividing the number of correct words by the total number of words used in the test multiplied by 100 [31]. The best accuracy results came from the proposed framework, followed by the B-Spline method for all lower resolutions used, 50 dpi, 75 dpi, 100 dpi, and 150 dpi. Tab. 1 also shows that the accuracy of the Bi-linear method looks close to the Bi-cubic method while the nearest neighbor got the worst accuracy results. Evaluation experiments prove that the proposed framework is the best compared to state-of-the-art interpolation algorithms.

5 Conclusion

This study analyzed the behavior of a text retrieval system when dealing with several low-resolution values. It is observed that the accuracy of text retrieval decreases significantly when the image resolution in the input documents becomes lower. To solve this challenge, existing interpolation algorithms, such as Nearest Neighbor, Bi-linear, Bi-cubic, and B-Spline, perform image re-scaling to obtain better quality. However, this study demonstrated that the combination features of multiple interpolation algorithms can lead to better results. The proposed framework can recover and maintain lost text details that may be lost during denoising with existing methods. The proposed framework was tested under four low-resolution documents of 50 dpi, 75 dpi, 100 dpi, and 150 dpi.

The experiment found that the proposed framework is the most efficient with a relative improvement of 11% compared to the best accuracy of existing methods. Hence, the preprocessing stage of the text retrieval system can incorporate the proposed framework to further improve the recognition accuracy. To further improve the proposed framework, a language model that preserves context information of statements can be added for a better selection of features among the interpolation algorithms used. Further research can also be done to determine the behavior of this framework when applied to images with different font types and sizes.

References

[1] Fernandes E., Moro S., Cortez P.,*Data science, machine learning and big data in digital journalism: a survey of state-of-the-art, challenges and opportunities*, Expert Systems with Applications, 221.4 (2023), 119795.

[2] Mager A., Norocel O.C., Rogers R.,*Advancing search engine studies: The evolution of Google critique and intervention*, Big Data & Society, 10.2 (2023), 521-528.

[3] Alahakoon D., Nawaratne R., Xu Y., De Silva D., Sivarajah U., Gupta B., *Self-building artificial intelligence and machine learning to empower big data analytics in smart cities*, Information Systems Frontiers, 22.8 (2020), 1-20.

[4] Luan Y., Eisenstein J., Toutanova K., Collins M., *Sparse, dense, and attentional representations for text retrieval*, Transactions of the Association for Computational Linguistics, 9.2 (2021), 329-345.

[5] Martinez-Rodriguez J.L., Hogan A., Lopez-Arevalo I., *Information extraction meets the semantic web: a survey*, Semantic Web, 11.2 (2020), 255-335.

[6] Georgiou T., Liu Y., Chen W., Lew M., *A survey of traditional and deep learning-based feature descriptors for high dimensional data in computer vision*, International Journal of Multimedia Information Retrieval, 9.5 (2020), 135-170.

[7] Dastile X., Celik T., Potsane M., *Statistical and machine learning models in credit scoring: A systematic literature survey*, Applied Soft Computing, 91.1 (2020), 106263.

[8] Ramadan M.J., Habeeb I.Q., *A systematic review in recent trends of e-learning usability evaluation techniques*, Proceedings of the AIP Conference Proceedings, 2023.

[9] Habeeb I., Al-Zaydi Z., Abdulkhudhur H., *Selection technique for multiple outputs of optical character recognition*, Eurasian Journal of Mathematical and Computer Applications, 8.2 (2020), 41-51.

[10] Chen H. et al., *Real-world single image super-resolution: A brief review*, Information Fusion, 79.3 (2022), 124-145.

[11] Xu X., Chen H., Moreno-Noguer F., Jeni L.A., De la Torre F., *3D human pose, shape and texture from low-resolution images and videos*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 44.9 (2021), 4490-4504.

[12] Souibgui M.A., Kessentini Y., *De-gan: A conditional generative adversarial network for document enhancement*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 44.3 (2020), 1180-1191.

[13] Jiang J., Wang C., Liu X., Ma J., *Deep learning-based face super-resolution: A survey*, ACM Computing Surveys, 55.1 (2021), 1-36.

[14] Singh A., Porikli F., Ahuja N., *Super-resolving noisy images*, Proceedings of the IEEE conference on computer vision and pattern recognition, Columbus, OH, USA, 2014, 2846-2853.

[15] Lund W.B., *Ensemble Methods for Historical Machine-Printed Document Recognition*, PhD dissertation, Brigham Young University, Utah, USA, 2014.

[16] Patel C., Patel A., Patel D., *Optical character recognition by open source OCR tool tesseract: A case study*, International Journal of Computer Applications, 55.10 (2012), 50-56.

[17] Malik A., Sikka G., Verma H.K., *A reversible data hiding scheme for interpolated images based on pixel intensity range*, Multimedia Tools and Applications, 79.3 (2020), 18005-18031.

[18] Karim S.A.A., *Rational bi-quartic spline with six parameters for surface interpolation with application in image enlargement*, IEEE Access, 8.1 (2020), 115621-115633.

[19] Rabbi J., Ray N., Schubert M., Chowdhury S., Chao D.,*Small-object detection in remote sensing images with end-to-end edge-enhanced GAN and object detector network*, Remote Sensing, 12.9 (2020), 1432-1439.

[20] Hirahara D., Takaya E., Kadowaki M., Kobayashi Y., Ueda T.,*Effect of the pixel interpolation method for downsampling medical images on deep learning accuracy*, Journal of Computer and Communications, 9.11 (2021), 150-156.

[21] Triwijoyoa B.K. and Adila A.,*Analysis of medical image resizing using bicubic interpolation algorithm*, Jurnal Ilmu Komputer, 14.2 (2021), 20-29.

[22] Cho W., Kwon J., Kwon S., and Yoo J., *A comparative study on OCR using super-resolution for small fonts*, International journal of advanced smart convergence, 8.3 (2019), 95-101.

[23] Chen X., Jin L., Zhu Y., Luo C., Wang T.,*Text recognition in the wild: A survey*, ACM Computing Surveys, 54.2 (2021), 1-35.

[24] Ahlawat S., Choudhary A., Nayyar A., Singh S., Yoon B.,*Improved handwritten digit recognition using convolutional neural networks*, Sensors, 20.12 (2020), 3344.

[25] Luo C., Jin L., Sun Z.,*Moran: A multi-object rectified attention network for scene text recognition*, Pattern Recognition, 90.2 (2019), 109-118.

[26] Li P., Prieto L., Mery D., Flynn P.J.,*On low-resolution face recognition in the wild: Comparisons and new techniques*, IEEE Transactions on Information Forensics and Security, 14.8 (2019), 2000-2012.

[27] Boukhtache S., Blaysat B., Grediac M., Berry F.,*Alternatives to bicubic interpolation considering FPGA hardware resource consumption*, IEEE Transactions on Very Large Scale Integration Systems, 29.2 (2020), 247-258.

[28] Rukundo O., *Evaluation of rounding functions in nearest neighbor interpolation*, International Journal of Computational Methods, 18.8 (2021), 2150024.

[29] Alipour S., Mirzaee F., *An iterative algorithm for solving two dimensional nonlinear stochastic integral equations: A combined successive approximations method with bilinear spline interpolation*, Applied Mathematics and Computation, 371.2 (2020), 124947.

[30] Zachariadis O. et al., *Accelerating B-spline interpolation on GPUs: Application to medical image registration*, Computer methods and programs in biomedicine, 193.1 (2020), 105431.

[31] Habeeb I., Al-Zaydi Z., Abdulkhudhur H.,*Speech Recognition in Videos Using Features of Different Enhancement Filters*, Eurasian Journal of Mathematical and Computer Applications, 12.3 (2024), 51-60.

Habeeb I.Q.
College of Biomedical Informatics,
University of Information Technology
and Communications,
Baghdad, Iraq,
Email: emadkassam@uoitc.edu.iq

Al-Zaydi Z.Q.
Biomedical Engineering, University of Technology,
Baghdad, Iraq,
Email: zeyad.q.habeeb@uotechnology.edu.iq

Abdulkhudhur H.N.
Directorate of Second Karkh, Ministry of Education,
Baghdad, Iraq,
Email: hanan_nagem@yahoo.com

Received 28.11.2024, Accepted 04.03.2025, Available online 31.12.2025.