

(12) PATENT APPLICATION PUBLICATION

(21) Application No.202311060844 A

(19) INDIA

(22) Date of filing of Application :11/09/2023

(43) Publication Date : 13/10/2023

(54) Title of the invention : A SYSTEM AND METHOD FOR SEGMENTING TEXT LINES IN A DOCUMENT

(51) International classification :G06T0007110000, G06T0007000000, G06T0011600000, G06T0005000000, H04L0045000000

(86) International Application No :NA
Filing Date :NA

(87) International Publication No : NA

(61) Patent of Addition to Application Number :NA
Filing Date :NA

(62) Divisional to Application Number :NA
Filing Date :NA

(71)Name of Applicant :

1)Chitkara University

Address of Applicant :Chitkara University, Chandigarh-Patiala National Highway, Village Jhansla, Rajpura, Punjab - 140401, India. Patiala -----

2)Bluest Mettle Solutions Private Limited

Name of Applicant : NA

Address of Applicant : NA

(72)Name of Inventor :

1)MISHRA, Rahul

Address of Applicant :ODC-4, Panchshil Tech Park, inside Courtyard by Marriott premises, Hinjewadi Phase - 1, Pune - 411057, Maharashtra, India. Pune -----

2)PANDEY, Sakshi

Address of Applicant :ODC-4, Panchshil Tech Park, inside Courtyard by Marriott premises, Hinjewadi Phase - 1, Pune - 411057, Maharashtra, India Pune -----

3)MANTRI, Archana

Address of Applicant :Chitkara University, Chandigarh-Patiala National Highway, Village Jhansla, Rajpura, Punjab - 140401, India Patiala -----

(57) Abstract :

Embodiments of the present disclosure relates to a system (100) and method (300) for segmenting text lines in documents. In an aspect, the present disclosure discloses a system (102) for segmenting text lines in documents using advanced image processing and machine learning techniques. The system (102) comprises a processor (202) coupled to a memory (204). The memory (204) stores processor-executable instructions. The processor (202) is configured to receive a document image from a user. Further, the processor (202) is configured to process the received document image. Next, the processor (202) is configured to segment the plurality of text lines in the processed document image. In the end, the processor (202) is configured to optimize the plurality of text lines in the document image based on the segmentation

No. of Pages : 27 No. of Claims : 10