

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property
Organization
International Bureau

(43) International Publication Date
10 September 2021 (10.09.2021)



(10) International Publication Number
WO 2021/176281 A1

(51) International Patent Classification:
G06T 1/00 (2006.01) *G06K 7/10* (2006.01)

(21) International Application Number:
PCT/IB2021/050612

(22) International Filing Date:
27 January 2021 (27.01.2021)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
16/811,490 06 March 2020 (06.03.2020) US

(71) Applicant: **INTERNATIONAL BUSINESS MACHINES CORPORATION** [US/US]; New Orchard Road, Armonk, New York, New York 10504 (US).

(71) Applicants (for MG only): **IBM ISRAEL - SCIENCE & TECHNOLOGY LTD.** [IL/IL]; Haifa University Campus, Mount Carmel, 31905 Haifa (IL). **IBM (CHINA) IN-**

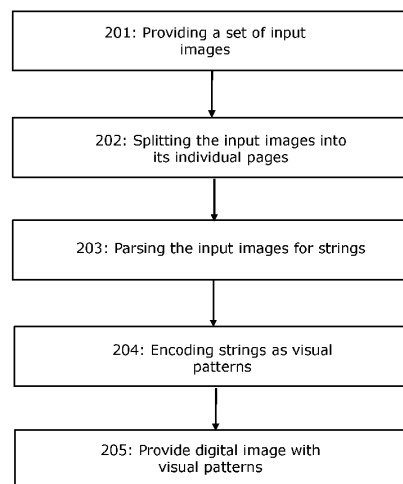
VESTMENT COMPANY LTD. [CN/CN]; 25/F, Pangu Plaza, No. 27, Central North 4th Ring Road, Chaoyang District, Beijing 100101 (CN).

(72) Inventors: **STAAR, Peter**; IBM Research GmbH, Saemmerstrasse 4, Rueschlikon, 8803 Zurich (CH). **DOLFI, Michele**; IBM Research GmbH, Saemmerstrasse 4, Rueschlikon, 8803 Zurich (CH). **AUER, Christoph**; IBM Research GmbH, Saemmerstrasse 4, Rueschlikon, 8803 Zurich (CH). **GEORGOPOULOS, Leonidas**; IBM Research GmbH, Saemmerstrasse 4, Rueschlikon, 8803 Zurich (CH). **BEKAS, Konstantinos**; IBM Research GmbH, Saemmerstrasse 4, Rueschlikon, 8803 Zurich (CH).

(74) Agent: **GILBOA, Eyal**; IBM Israel - Science & technology Ltd., Haifa University Campus, Mount Carmel, 31905 Haifa (IL).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ,

(54) Title: DIGITAL IMAGE PROCESSING



Encode NL features

FIG. 2a

(57) Abstract: A computer-implemented method for processing a digital image. The digital image comprises one or more text cells, wherein each of the one or more text cells comprises a string and a bounding box. The method comprises receiving the digital image in a first format, the first format providing access to the strings and the bounding boxes of the one or more text cells. The method further comprises encoding the strings of the one or more text cells as visual pattern according to a predefined string encoding scheme and providing the digital image in a second format. The second format comprises the visual pattern of the strings of the one or more text cells. A corresponding system and a related computer program product is provided.

[Continued on next page]



WO 2021/176281 A1

CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO,
DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN,
HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN,
KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD,
ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO,
NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW,
SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN,
TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

- with international search report (Art. 21(3))
- in black and white; the international application as filed contained color or greyscale and is available for download from PATENTSCOPE