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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
03 September 2020 (03.09.2020)

WIPO | PCT

(10) International Publication Number
WO 2020/174447 A1

(51) International Patent Classification:
G06T 11/00 (2006.01) *G06T 7/00* (2017.01)

(21) International Application Number:
PCT/IB2020/051706

(22) International Filing Date:
28 February 2020 (28.02.2020)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
201921007962 28 February 2019 (28.02.2019) IN

(71) Applicant: **TATA CONSULTANCY SERVICES LIMITED** [IN/IN]; Nirmal Building, 9th Floor, Nariman Point, Mumbai 400021 (IN).

(72) Inventors: **HEBBALAGUPPE, Ramya Sugnana Murthy**; Tata Consultancy Services Limited, Block C, Kings Canyon, ASF Insignia, Faridabad Road, Gawal Pahari, Gurgaon 122003, Haryana (IN). **HEGDE, Srinidhi**; Tata Consultancy Services Limited, Block C, Kings Canyon, ASF Insignia, Faridabad Road, Gawal Pahari, Gurgaon 122003, Haryana (IN). **MAURYA, Jitender Kumar**; Tata Consultancy Services Limited, Block C, Kings Canyon, ASF Insignia, Faridabad Road, Gawal Pahari, Gurgaon 122003, Haryana (IN).

(74) Agent: **KHAITAN & CO**; One Indiabulls Centre, 13th Floor, 841, Senapati Bapat Marg, Elphinstone Road, Mumbai 400013, Maharashtra (IN).

(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, 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, 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,

(54) Title: MULTI-LABEL PLACEMENT FOR AUGMENTED AND VIRTUAL REALITY AND VIDEO ANNOTATIONS

(57) Abstract: Typically when labels are randomly fused into a video, it results in occlusion of main subjects in every video frames. Further, random placement of labels corresponding to multiple objects in the frame may confuse the user as he/she may struggle to identify label corresponding to each object. Disclosed herein are a method and system for identifying optimum location for label placement in a video. For a given video, the system generates a plurality of object-label pairs, and also a saliency map. The object-label pairs and the saliency map are processed by the system to identify the optimum location for placing each label such that at the optimum location conditions related to occlusion, closeness to object, intersection between connectors, and diagonal heuristic and central bias are satisfied.

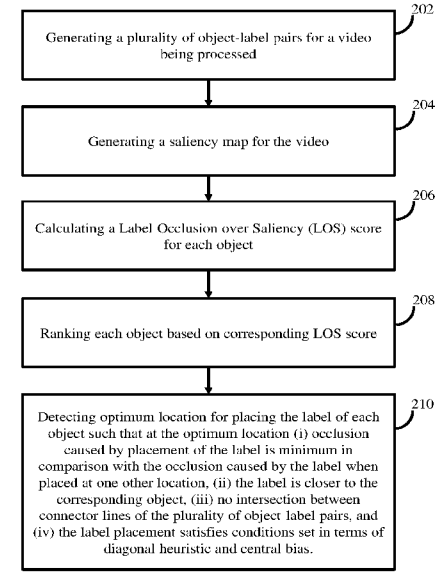


Fig. 2

WO 2020/174447 A1

[Continued on next page]

TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

Published:

— *with international search report (Art. 21(3))*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

— *in black and white; the international application as filed contained color or greyscale and is available for download from PATENTSCOPE*

MULTI-LABEL PLACEMENT FOR AUGMENTED AND VIRTUAL REALITY AND VIDEO ANNOTATIONS

CROSS-REFERENCE TO RELATED APPLICATIONS AND PRIORITY

[001] The present PCT application claims priority to India Patent Application No. 201921007962, filed before Indian Patent Office on February 28, 2019. Entire contents of the aforementioned application are incorporated herein by reference.

TECHNICAL FIELD

[002] The disclosure herein generally relates to video processing, and, more particularly, to a method and system for finding optimum location for label placement in a video.

BACKGROUND

[003] In various applications such as but not limited to augmented/virtual reality, objects in a video are labelled for the benefit of users. This helps the users understand what/who each object is, along with any additional information. In the augmented reality based applications, fusion of contextual synthetic data with the visual data (video) enriches perception and efficiency of a user who is performing a task. Contextual data (for example, labels, coordinates, and so on) that are inserted to the video are called overlays.

[004] Size and shape of such overlays may vary from one to other. When such overlays are fused to the visual data, it is possible that the overlays may cause occlusion of actual objects in the video. In addition to this, consider a scenario in which in a particular frame in the video multiple objects are present. If labels corresponding to all the objects are placed randomly, the user may find it confusing to understand labels matching each of the objects.

SUMMARY

[005] Embodiments of the present disclosure present technological improvements as solutions to one or more of the above-mentioned technical problems recognized by the inventors in conventional systems. For example, in one embodiment, a processor implemented method for label placement in a video is provided. In this method, the video is collected as input, by one or more hardware processors. Further, a plurality of object-label pairs are generated for the video, by the one or more hardware processors. Then a saliency map is generated for the video, by the one or more hardware processors, wherein the saliency map indicates saliency of a plurality of regions in each frame of the video. Further an optimum location for placing the label of each of the object-label pairs is detected. In the process of detecting the optimum location, a Label Occlusion over Saliency (LOS) score for each of the objects is calculated, and then each of the plurality of object-label pairs is ranked based on corresponding LOS score. Further, the optimum location is detected such that at the optimum location (i) occlusion caused by placement of the label is minimum in comparison with the occlusion caused by the label when placed at one other location, (ii) the label is closer to the corresponding object, (iii) no intersection between connector lines of the plurality of object-label pairs, and (iv) the label placement satisfies conditions set in terms of diagonal heuristic and central bias.

[006] In another aspect, a system for label placement in a video is provided. The system includes a memory module storing a plurality of instructions; one or more communication interfaces; and one or more hardware processors coupled to the memory module via the one or more communication interfaces. The one or more hardware processors are caused by the plurality of instructions to collect the video is collected as input. Further, a plurality of object-label pairs are generated for the video. Then a saliency map is generated for the video, wherein the saliency map indicates saliency of a plurality of regions in each frame of the video. Further an optimum location for placing the label of each of the object-label pairs is detected. In the process of detecting the optimum location, a Label Occlusion over Saliency (LOS) score for each of the objects is calculated, and then

each of the plurality of object-label pairs is ranked based on corresponding LOS score. Further, the optimum location is detected such that at the optimum location (i) occlusion caused by placement of the label is minimum in comparison with the occlusion caused by the label when placed at one other location, (ii) the label is closer to the corresponding object, (iii) no intersection between connector lines of the plurality of object-label pairs, and (iv) the label placement satisfies conditions set in terms of diagonal heuristic and central bias.

[007] In yet another aspect, a non-transitory computer readable medium for label placement in a video is provided. The non-transitory computer readable medium executes the following method to identify an optimum location for label placement. In this method, the video is collected as input, by one or more hardware processors. Further, a plurality of object-label pairs are generated for the video, by the one or more hardware processors. Then a saliency map is generated for the video, by the one or more hardware processors, wherein the saliency map indicates saliency of a plurality of regions in each frame of the video. Further an optimum location for placing the label of each of the object-label pairs is detected. In the process of detecting the optimum location, a Label Occlusion over Saliency (LOS) score for each of the objects is calculated, and then each of the plurality of object-label pairs is ranked based on corresponding LOS score. Further, the optimum location is detected such that at the optimum location (i) occlusion caused by placement of the label is minimum in comparison with the occlusion caused by the label when placed at one other location, (ii) the label is closer to the corresponding object, (iii) no intersection between connector lines of the plurality of object-label pairs, and (iv) the label placement satisfies conditions set in terms of diagonal heuristic and central bias.

[008] It is to be understood that both the foregoing general description and the following detailed description are exemplary and explanatory only and are not restrictive of the invention, as claimed.

BRIEF DESCRIPTION OF THE DRAWINGS

[009] The accompanying drawings, which are incorporated in and

constitute a part of this disclosure, illustrate exemplary embodiments and, together with the description, serve to explain the disclosed principles:

[010] FIG. 1 illustrates an exemplary block diagram of a system for determining optimum location for label placement, according to some embodiments of the present disclosure.

[011] FIG. 2 is a flow diagram depicting steps involved in the process of determining optimum location for label placement, using the system of FIG. 1, according to some embodiments of the present disclosure.

[012] FIG. 3 is an example diagram depicting data and data flow in the process of determining optimum location for label placement being performed using the system of FIG. 1, according to some embodiments of the present disclosure.

[013] FIG. 4 (a through e) are example diagrams depicting different properties considered for diagonal heuristic and the central bias, according to some embodiments of the present disclosure.

[014] FIG. 5 (a through e) are example diagrams depicting involved in the process of determining optimum location for label placement, using the system of FIG. 1, according to some embodiments of the present disclosure.

DETAILED DESCRIPTION OF EMBODIMENTS

[015] Exemplary embodiments are described with reference to the accompanying drawings. In the figures, the left-most digit(s) of a reference number identifies the figure in which the reference number first appears. Wherever convenient, the same reference numbers are used throughout the drawings to refer to the same or like parts. While examples and features of disclosed principles are described herein, modifications, adaptations, and other implementations are possible without departing from the spirit and scope of the disclosed embodiments. It is intended that the following detailed description be considered as exemplary only, with the true scope and spirit being indicated by the following claims.

[016] Referring now to the drawings, and more particularly to FIG. 1 through FIG. 5, where similar reference characters denote corresponding features

consistently throughout the figures, there are shown preferred embodiments and these embodiments are described in the context of the following exemplary system and/or method.

[017] FIG. 1 illustrates an exemplary block diagram of a system for determining optimum location for label placement, according to some embodiments of the present disclosure. The system 100 includes at least one memory module 101, at least one hardware processor 102, and at least one communication interface 103.

[018] The one or more hardware processors 102 can be implemented as one or more microprocessors, microcomputers, microcontrollers, digital signal processors, central processing units, state machines, graphics controllers, logic circuitries, and/or any devices that manipulate signals based on operational instructions. Among other capabilities, the hardware processor(s) 102 are configured to fetch and execute computer-readable instructions stored in the memory module 101, which causes the hardware processor(s) 102 to perform actions depicted in FIG. 2 for the purpose of identifying the optimum location for label placement. In an embodiment, the system 100 can be implemented in a variety of computing systems, such as laptop computers, notebooks, hand-held devices, workstations, mainframe computers, servers, a network cloud and the like.

[019] The communication interface(s) 103 can include a variety of software and hardware interfaces, for example, a web interface, a graphical user interface, and the like and can facilitate multiple communications within a wide variety of networks N/W and protocol types, including wired networks, for example, LAN, cable, etc., and wireless networks, such as WLAN, cellular, or satellite. In an embodiment, the communication interface(s) 103 can include one or more ports for connecting a number of devices to one another or to another server.

[020] The memory module(s) 101 may include any computer-readable medium known in the art including, for example, volatile memory, such as static random access memory (SRAM) and dynamic random access memory (DRAM), and/or non-volatile memory, such as read only memory (ROM), erasable programmable ROM, flash memories, hard disks, optical disks, and magnetic tapes. In an embodiment, one or more modules (not shown) of the system 100 can be

stored in the memory 101. The memory module(1) 101 stores a plurality of instructions which when executed, cause the one or more hardware processors 102 to perform one or more actions and corresponding the identification of optimum location for label placement being handled by the system 100.

[021] The system 100 collects a video for processing. In an embodiment the system 100 may collect and process more than one video at a time. The video may be an RGB video V <f_1, f_2, ..., f_n> with a frame sequence of length 'n' and each frame having a dimension of F_w * F_h.

[022] The system 100 then uses/executes any suitable mechanism/technique to process each frame of the video, one frame at a time or multiple frames at a time, to identify one or more objects in each frame. In an embodiment, all the identified objects are labelled by the system 100. In another embodiment, out of a plurality of objects identified, at least one object is selected as an Object of Interest and then only the selected at least one object of interest is labelled. The system 100 uses appropriate mechanism(s) to generate at least one label for each object. For example, YOLOv2 mechanism may be used by the system 100 for identifying and labeling the objects. The objects and the corresponding labels are used by the system 100 to generate a plurality of object-label pairs corresponding to the video being processed.

[023] The system 100 then uses a Saliency Attention Model (SAM) for generating a saliency map corresponding to the video being processed. SAM predicts saliency of regions in each frame being processed, and this information is captured in the saliency map. In addition to the saliency information, the saliency map may also include data pertaining to identified eye fixation points of the user on each frame of the video.

[024] The object-label pairs and the saliency map(s) are then processed further by the system 100 to identify the optimum location (represented in terms of coordinates of the location) for placing each of the generated labels. At this stage, the system 100 considers each object sequentially, in decreasing order of saliency occlusion (as indicated in the saliency map). Every time an overlay (which may be the label or any other type of overlay) is placed, the corresponding region (i.e. the

region occupied by the overlay) is marked as highly salient region, which in turn indicates that this region is not suitable for placing another overlay.

[025] The system 100 then calculates a Label Occlusion over Saliency (LOS) score of bounding box of each object being considered, wherein the LOS score of an object represents saliency occlusion by the object and corresponding label. The LOS score is calculated as:

$$\text{LOS (N, G)} = \frac{\sum_{(x,y)\in N} G(x,y)}{|N|} \text{ --- (1)}$$

Where N is a set of pixels (x, y) that is occluded by the overlay and G is a ground truth saliency map. The LOS score ranges from 0 to 1, where score of 0 represents no occlusion with any salient region and score of 1 represents complete overlap with the high salient region.

[026] The system 100 places labels such that the system 100 avoids placement of the labels on the objects and previously placed labels. In addition to minimizing occlusion, the system 100 requires the optimum location (and the coordinates) to satisfy three other conditions, namely 1. Closeness of the label to corresponding object, 2. No/minimal intersection between connector lines of the plurality of object-label pairs, and 3. Conditions set in terms of diagonal heuristic and central bias.

[027] The system 100 checks and verifies conditions in terms of the closeness of the labels to objects and the intersection between the connector lines using Voronoi partitioning of each frame being processed. The system 100 performs Voronoi partitioning of each of the frames, by keeping centroids of the bounding boxes of the objects as seeding points. The Voronoi partitioning divides each frame to a plurality of regions such that each object in the frame is encompassed in a corresponding region. By keeping centroids of each of the bounding boxes as the seed point for corresponding region, the system 100 is able to ensure that top left corner of a label is placed close to the corresponding object.

[028] The system 100 further uses the Voronoi partitioning to ensure minimal/no intersection between the connector lines. Connector lines are the lead lines that connect an object to corresponding label. As the system 100 uses the

Voronoi partitioning data, the start and end points of each connector lines may be selected such that the start and end points of a connector line of an object remains within region of that object. As a result of this approach, Euclidean distance between top left corner of the label and the centroid of a bounding box of the corresponding object is minimum. As each object is within separate regions in the Voronoi partitioning, this approach ensures that the connector lines do not intersect, which in turn improves user experience. Given below is a proof that the intersection between the connector lines can be removed using this approach:

[029] Let ρ1 and ρ2 be object bounding box centroids, which are also the seed points for the respective Voronoi partitions, V1 and V2. Consider two distinct connectors, C (ρ_1,ρ_1') between endpoints $\rho_1$andρ_1', and C (ρ_2,ρ_2') between endpoints ρ_2 andρ_2'. Voronoi partitions are convex polygons. From the definition of convexity, all the points s on the line segment C(s_1, s_2) also lie in the corresponding Voronoi region, i.e., if ρ lies on the line segment C(ρ_1, ρ_1'), then it also lies within V_1. Assume that C (ρ_1,ρ_1') and C (ρ_2,ρ_2') intersect at x, which implies that x ε V_1 ∩ V_2. For a strict Voronoi partition, $V_1 \cap V_2 = \emptyset$, hence, the connectors are the same. However this leads to a contradiction since C (ρ_1,ρ_1') and C (ρ_2,ρ_2') are distinct. Thus C (ρ_1,ρ_1') and C (ρ_2,ρ_2') never intersect.

[030] The system further ensures that the optimum location satisfies the condition in terms of the diagonal heuristic and the central bias. Studies have indicates that placing labels on diagonal angle bisectors improves user experience, and that eye-fixation points tend to cluster towards centre of the screen, a property of the human eyes termed as the 'central bias'. These properties are depicted in FIG. 4.

[031] The system 100 outputs the optimum location (and corresponding coordinates in the frames), such that the label placement at these coordinates satisfies the aforementioned conditions, and in turn improves user experience. The system 100 follows the aforementioned approach so as to place multiple labels (multi label placement) within a video, as part of annotating the video (or annotating the objects in the video), in applications such as but not limited to augmented reality/virtual reality.

[032] FIG. 2 is a flow diagram depicting steps involved in the process of determining optimum location for label placement, using the system of FIG. 1, according to some embodiments of the present disclosure. The system 100 collects a video for processing, and by processing the collected video, generates (202) a plurality of object-label pairs corresponding to each frame in the video. The system 100 then generates (204) at least one saliency map for the video, wherein the saliency map indicates saliency of regions in each frame being considered.

[033] The system 100 then calculates (206) LOS score for each object being considered, and then each object is ranked (208) based on the corresponding LOS score. The system 100 then determines (210) optimum location for placing each label (corresponding to each object) such that, at the optimum location, (i) occlusion caused by placement of the label is minimum in comparison with the occlusion caused by the label when placed at any other location, (ii) the label is closer to the corresponding object, (iii) no intersection between connector lines of the plurality of object-label pairs, and (iv) the label placement satisfies conditions set in terms of diagonal heuristic and central bias.

[034] The optimum location(s) thus identified and the corresponding coordinates are then provided as output by the system 100. Data flow in this mechanism is depicted in FIG. 3 as well. Further, the different steps involved in the process of identifying the optimum location for label placement are schematically represented in FIG. 5.

Experimental Results:

[035] Deep learning models for object detection and SAM were trained in PyTorch. For object detection and label generation, YOLOv2 pre-trained on COCO dataset having 80 classes was used. Input video size was resized to 608*608 resolution before feeding as input to YOLOv2. The SAM that was used for computing saliency maps had been pre-trained on SALICON dataset containing eye fixation ground truth for images.

1) Saliency map computation:

[036] During the experiment conducted, accuracy of the saliency prediction being carried out by the system 100 was compared with multiple baseline

methods such as NSS, CC, AUC (Judd), sAUC, and KL. Saliency evaluation was carried out on SALICON dataset, and results are shown in Table.1.

Methods	NSS (↑)	CC (↑)	AUC-J (↑)	sAUC (↑)	KL (↓)
DeepFix	2.26	0.78	0.87	0.71	0.63
SaIGAN	2.04	0.73	0.86	0.72	1.07
SAM-Resnet	2.34	0.78	0.87	0.70	1.27
SAM-VGG	2.30	0.77	0.87	0.71	1.13

Table. 1

[037] Mean of three Gaussian Priors $\mathcal{N}(\mu_1, \sigma_1)$, $\mathcal{N}(\mu_2, \sigma_2)$, $\mathcal{N}(\mu_3, \sigma_3)$ were used for modelling the central bias. Here $\mu_1 = \mu_2 = \mu_3 = (0.5 * F_w, 0.5 * F_h)$, $\sigma_1 = (0.5 * \min(F_w, F_h), (0.5 * \min(F_w, F_h))$, $\sigma_2 = (0.75 * \min(F_w, F_h), (0.25 * \min(F_w, F_h))$, and $\sigma_3 = (0.25 * \min(F_w, F_h), (0.75 * \min(F_w, F_h))$. A weighted average of the saliency map, the central bias, and diagonal heuristic was performed. More weight (of 0.7) was given to the predicted saliency map, and less weight (of 0.3) was given to the mask. It was observed that using the predicted saliency map with the diagonal heuristics gave better LOS score in comparison with addition of central bias component with saliency map. This is evident from Table. 2.

Decay functions	Average LOS score (↓)
No mask	0.0147
Only central bias	3.6450
Only diagonal heuristic	0.0071
Central bias + Diagonal heuristic	4.3310

Table. 2 (Comparison of performances with biases with linear and

exponential decay for including diagonal heuristic and central bias)

2) Overlay location prediction:

[038] During the overlay prediction, in order to improve temporal consistency in label placement, label locations were computed after skipping k frames of the video. Experiments proved that keeping value of k as 20 for a 30 fps video gave best rating for temporal coherence.

User Evaluation:

[039] User evaluation was carried out to understand whether the users found overlay placement mechanism being claimed useful or not. 21 subjects in total were selected out of which 9 subjects belong to age group of 20-25, 4 subjects belong to age group of 26-30, 5 subjects belong to age group of 31-35, and 3 subjects belong to age group > 35. Out of the 21 subjects, 13 were male and 8 were female subjects.

[040] The subjects viewed 20 recorded videos with different video resolutions from DIEM dataset which contained labels placed using the proposed mechanism. This datasets consisted of varieties of videos from different genres of advertisements, trailers, television-series, with scenes varying from nature to animated cartoons. Also with eye movements, this dataset provides detailed eye fixation saliency annotations. The users were tasked to rate the following label placement objectives for each video on a rating scale ranging from 1 to 5, 5 being the highest rating. The label placement objectives are also the subjective metrics as follows: (1) Occlusion Avoidance: Does the label cover/overlap with the regions of interest? Here, a rating of 5 means no occlusion with the salient regions of the videos (2) Proximity: Is the label placed close to the corresponding object? A rating of 5 corresponds to the label being very close to the object of interest. (3)Temporal Coherence: Are the labels jittery or jumpy? A rating of 5 means seamless transitions of labels in videos. (4) Readability: Is the label readable in every frame? A rating of 5 corresponds to the highest ease with which one can read especially the color of overlay box and text. (5) Color Scheme: Does the label font color stand out with respect to the background? Here 5 means contrast between label and background is high. (6) Clarity: Do the connectors or the leader lines intersect? Answers could be

Yes/No only.

[041] These metrics captured evaluate (a) user experience and (b) placement of overlays. In all the experiments, label dimensions D/K were used, where D is image dimension and K ε {4, 8, 12, 32}. This could be customized as per the users' needs. The videos were shown on a desktop and a laptop. Thereafter, we capture the mean opinion ratings for each of the six metrics.

[042] The written description describes the subject matter herein to enable any person skilled in the art to make and use the embodiments. The scope of the subject matter embodiments is defined by the claims and may include other modifications that occur to those skilled in the art. Such other modifications are intended to be within the scope of the claims if they have similar elements that do not differ from the literal language of the claims or if they include equivalent elements with insubstantial differences from the literal language of the claims.

[043] The embodiments of present disclosure herein addresses unresolved problem of label placement in a video. The embodiment, thus provides a mechanism for identifying an optimum location for placing a label in a video being processed.

[044] It is to be understood that the scope of the protection is extended to such a program and in addition to a computer-readable means having a message therein; such computer-readable storage means contain program-code means for implementation of one or more steps of the method, when the program runs on a server or mobile device or any suitable programmable device. The hardware device can be any kind of device which can be programmed including e.g. any kind of computer like a server or a personal computer, or the like, or any combination thereof. The device may also include means which could be e.g. hardware means like e.g. an application-specific integrated circuit (ASIC), a field-programmable gate array (FPGA), or a combination of hardware and software means, e.g. an ASIC and an FPGA, or at least one microprocessor and at least one memory with software modules located therein. Thus, the means can include both hardware means and software means. The method embodiments described herein could be implemented in hardware and software. The device may also include software means. Alternatively, the embodiments may be implemented on different hardware

devices, e.g. using a plurality of CPUs.

[045] The embodiments herein can comprise hardware and software elements. The embodiments that are implemented in software include but are not limited to, firmware, resident software, microcode, etc. The functions performed by various modules described herein may be implemented in other modules or combinations of other modules. For the purposes of this description, a computer-usable or computer readable medium can be any apparatus that can comprise, store, communicate, propagate, or transport the program for use by or in connection with the instruction execution system, apparatus, or device.

[046] The illustrated steps are set out to explain the exemplary embodiments shown, and it should be anticipated that ongoing technological development will change the manner in which particular functions are performed. These examples are presented herein for purposes of illustration, and not limitation. Further, the boundaries of the functional building blocks have been arbitrarily defined herein for the convenience of the description. Alternative boundaries can be defined so long as the specified functions and relationships thereof are appropriately performed. Alternatives (including equivalents, extensions, variations, deviations, etc., of those described herein) will be apparent to persons skilled in the relevant art(s) based on the teachings contained herein. Such alternatives fall within the scope and spirit of the disclosed embodiments. Also, the words "comprising," "having," "containing," and "including," and other similar forms are intended to be equivalent in meaning and be open ended in that an item or items following any one of these words is not meant to be an exhaustive listing of such item or items, or meant to be limited to only the listed item or items. It must also be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural references unless the context clearly dictates otherwise.

[047] Furthermore, one or more computer-readable storage media may be utilized in implementing embodiments consistent with the present disclosure. A computer-readable storage medium refers to any type of physical memory on which information or data readable by a processor may be stored. Thus, a computer-

readable storage medium may store instructions for execution by one or more processors, including instructions for causing the processor(s) to perform steps or stages consistent with the embodiments described herein. The term "computer-readable medium" should be understood to include tangible items and exclude carrier waves and transient signals, i.e., be non-transitory. Examples include random access memory (RAM), read-only memory (ROM), volatile memory, nonvolatile memory, hard drives, CD ROMs, DVDs, flash drives, disks, and any other known physical storage media.

[048] It is intended that the disclosure and examples be considered as exemplary only, with a true scope and spirit of disclosed embodiments being indicated by the following claims.

CLAIMS

1. A processor implemented method for label placement in a video, comprising:

collecting the video as input, by one or more hardware processors;

generating a plurality of object-label pairs for the video, by the one or more hardware processors;

generating a saliency map for the video, by the one or more hardware processors, wherein the saliency map indicates saliency of a plurality of regions in each frame of the video; and

detecting an optimum location for placing the label of each of the object-label pairs, comprising:

calculating a Label Occlusion over Saliency (LOS) score for each of the objects;

ranking each of the plurality of object-label pairs based on corresponding LOS score; and

detecting the optimum location such that at the optimum location (i) occlusion caused by placement of the label is minimum in comparison with the occlusion caused by the label when placed at any other location, (ii) the label is closer to the corresponding object, (iii) no intersection between connector lines of the plurality of object-label pairs, and (iv) the label placement satisfies conditions set in terms of diagonal heuristic and central bias.

2. The method as claimed in claim 1, wherein the occlusion caused by placement of the label for an object-label pair is determined based on the LOS score of the object in the object-label pair.

3. The method as claimed in claim 1, wherein closeness of the label to the corresponding object is determined based on Voronoi partitioning, such that

at the optimum location top left corner of the label is close to the corresponding object.

4. The method as claimed in claim 1, wherein the intersection between the connector lines of the plurality of object-label pairs is avoided based on Voronoi partitioning, such that at the optimum location (i) Euclidean distance between top left corner of the label and the centroid of a bounding box of the corresponding object is minimum, and (ii) the connector line of the object stays within same Voronoi partition as that of the centroid of the bounding box of the object.

5. A system for label placement in a video, comprising:
a memory module (101) storing a plurality of instructions;
one or more communication interfaces (103); and
one or more hardware processors (102) coupled to the memory module (101) via the one or more communication interfaces (103), wherein the one or more hardware processors are caused by the plurality of instructions to:
collect the video as input;
generate a plurality of object-label pairs for the video;
generate a saliency map for the video, wherein the saliency map indicates saliency of a plurality of regions in each frame of the video; and
detect an optimum location for placing the label of each of the object-label pairs, by:
calculating a Label Occlusion over Saliency (LOS) score for each of the objects;
ranking each of the plurality of object-label pairs based on corresponding LOS score; and
detecting the optimum location such that at the optimum location (i) occlusion caused by placement of the label is minimum in comparison with the occlusion caused by the

label when placed at one other location, (ii) the label is closer to the corresponding object, (iii) no intersection between connector lines of the plurality of object-label pairs, and (iv) the label placement satisfies conditions set in terms of diagonal heuristic and central bias.

6. The system as claimed in claim 5, wherein the occlusion caused by placement of the label for an object-label pair is determined based on the LOS score of the object in the object-label pair.

7. The system as claimed in claim 5, wherein the system determines the closeness of the label to the corresponding object based on Voronoi partitioning, such that at the optimum location top left corner of the label is close to the corresponding object.

8. The system as claimed in claim 5, wherein the system avoids intersection between the connector lines of the plurality of object-label pairs based on Voronoi partitioning, such that at the optimum location (i) Euclidean distance between top left corner of the label and the centroid of a bounding box of the corresponding object is minimum, and (ii) the connector line of the object stays within same Voronoi partition as that of the centroid of the bounding box of the object.

9. A non-transitory computer readable medium for label placement in a video, the non-transitory computer readable medium performs the label placement in the video by:

collecting the video as input, by one or more hardware processors;
generating a plurality of object-label pairs for the video, by the one or more hardware processors;

generating a saliency map for the video, by the one or more hardware processors, wherein the saliency map indicates saliency of a plurality of regions in each frame of the video; and

detecting an optimum location for placing the label of each of the object-label pairs, comprising:

calculating a Label Occlusion over Saliency (LOS) score for each of the objects;

ranking each of the plurality of object-label pairs based on corresponding LOS score; and

detecting the optimum location such that at the optimum location (i) occlusion caused by placement of the label is minimum in comparison with the occlusion caused by the label when placed at any other location, (ii) the label is closer to the corresponding object, (iii) no intersection between connector lines of the plurality of object-label pairs, and (iv) the label placement satisfies conditions set in terms of diagonal heuristic and central bias.

10. The non-transitory computer readable medium as claimed in claim 9, wherein the occlusion caused by placement of the label for an object-label pair is determined based on the LOS score of the object in the object-label pair.

11. The non-transitory computer readable medium as claimed in claim 9, wherein closeness of the label to the corresponding object is determined based on Voronoi partitioning, such that at the optimum location top left corner of the label is close to the corresponding object.

12. The non-transitory computer readable medium as claimed in claim 9, wherein the intersection between the connector lines of the plurality of object-label pairs is avoided based on Voronoi partitioning, such that at the

optimum location (i) Euclidean distance between top left corner of the label and the centroid of a bounding box of the corresponding object is minimum, and (ii) the connector line of the object stays within same Voronoi partition as that of the centroid of the bounding box of the object.

5

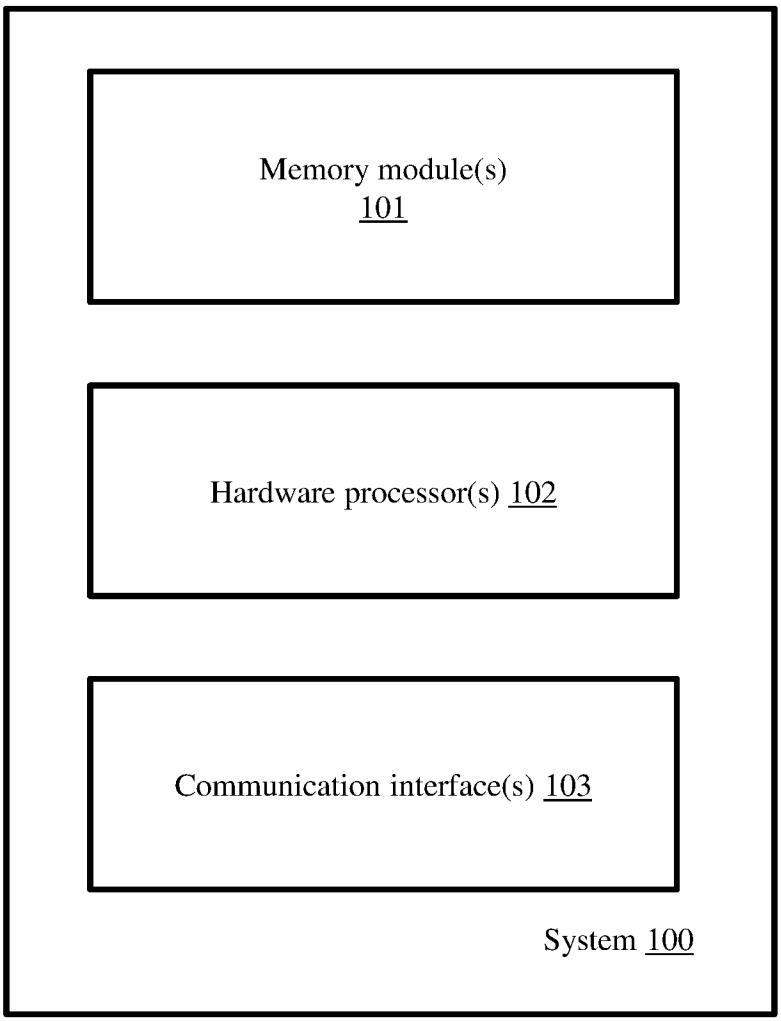

Memory module(s)
101
Hardware processor(s) 102
Communication interface(s) 103
System 100

Fig. 1

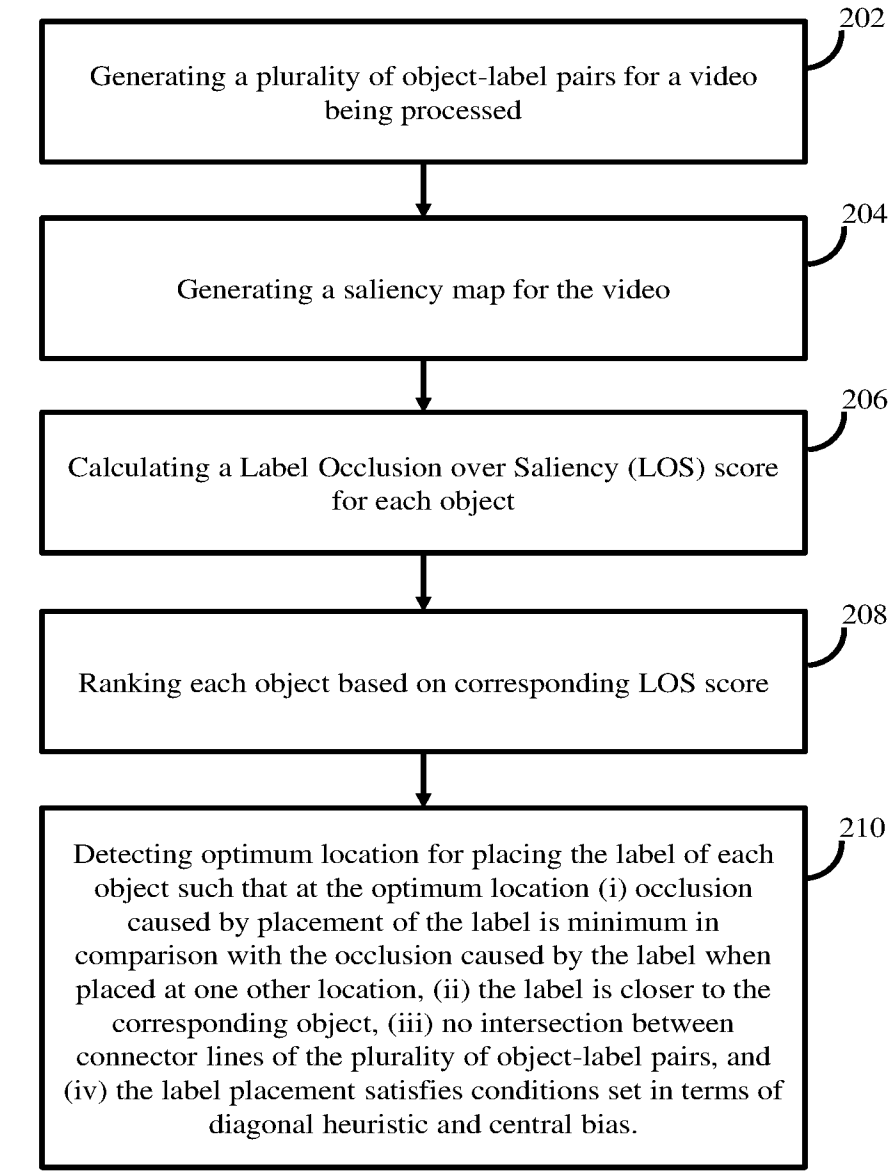

202
Generating a plurality of object-label pairs for a video being processed
204
Generating a saliency map for the video
206
Calculating a Label Occlusion over Saliency (LOS) score for each object
208
Ranking each object based on corresponding LOS score
210
Detecting optimum location for placing the label of each object such that at the optimum location (i) occlusion caused by placement of the label is minimum in comparison with the occlusion caused by the label when placed at one other location, (ii) the label is closer to the corresponding object, (iii) no intersection between connector lines of the plurality of object-label pairs, and (iv) the label placement satisfies conditions set in terms of diagonal heuristic and central bias.

Fig. 2

200

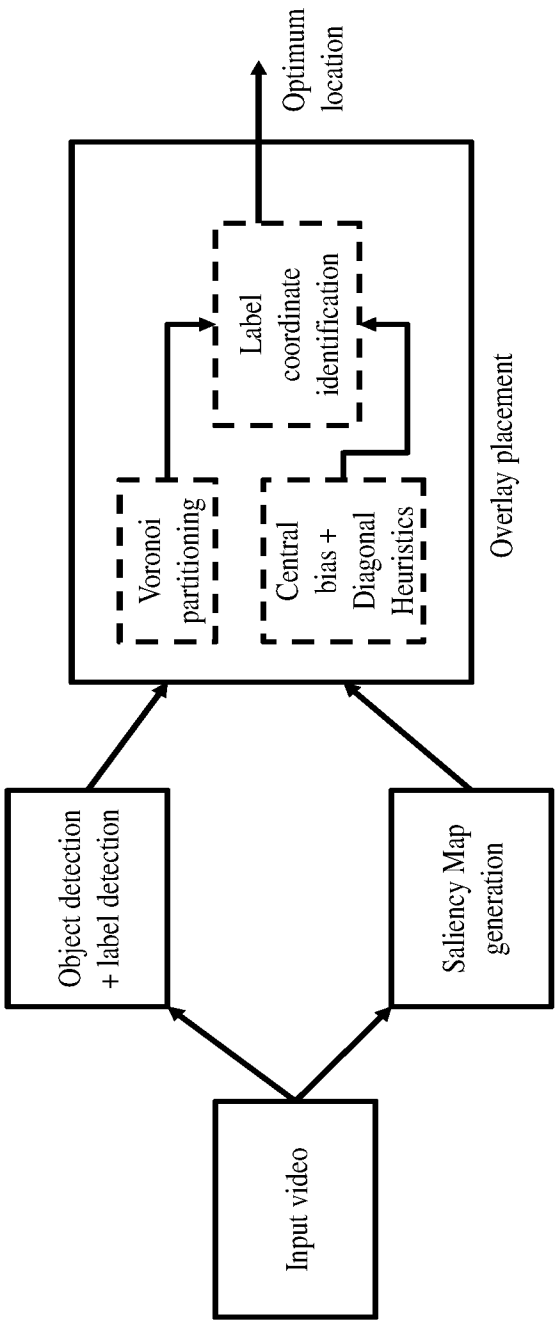
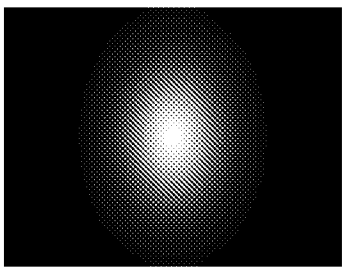
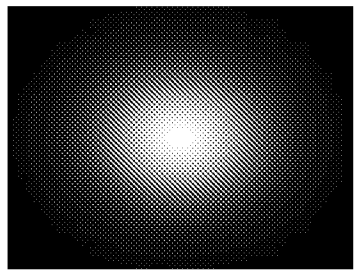

Input video
Object detection + label detection
Saliency Map generation
Overlay placement
Voronoi partitioning
Central bias + Diagonal Heuristics
Label coordinate identification
Optimum location

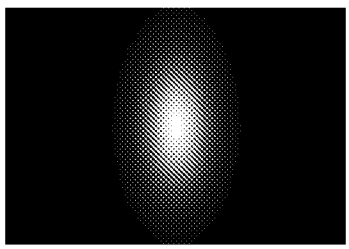
Fig. 3



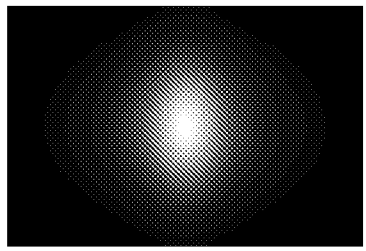
(a)



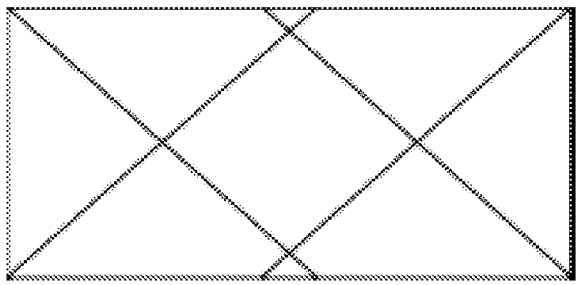
(b)



(c)



(d)



(e)

Fig. 4

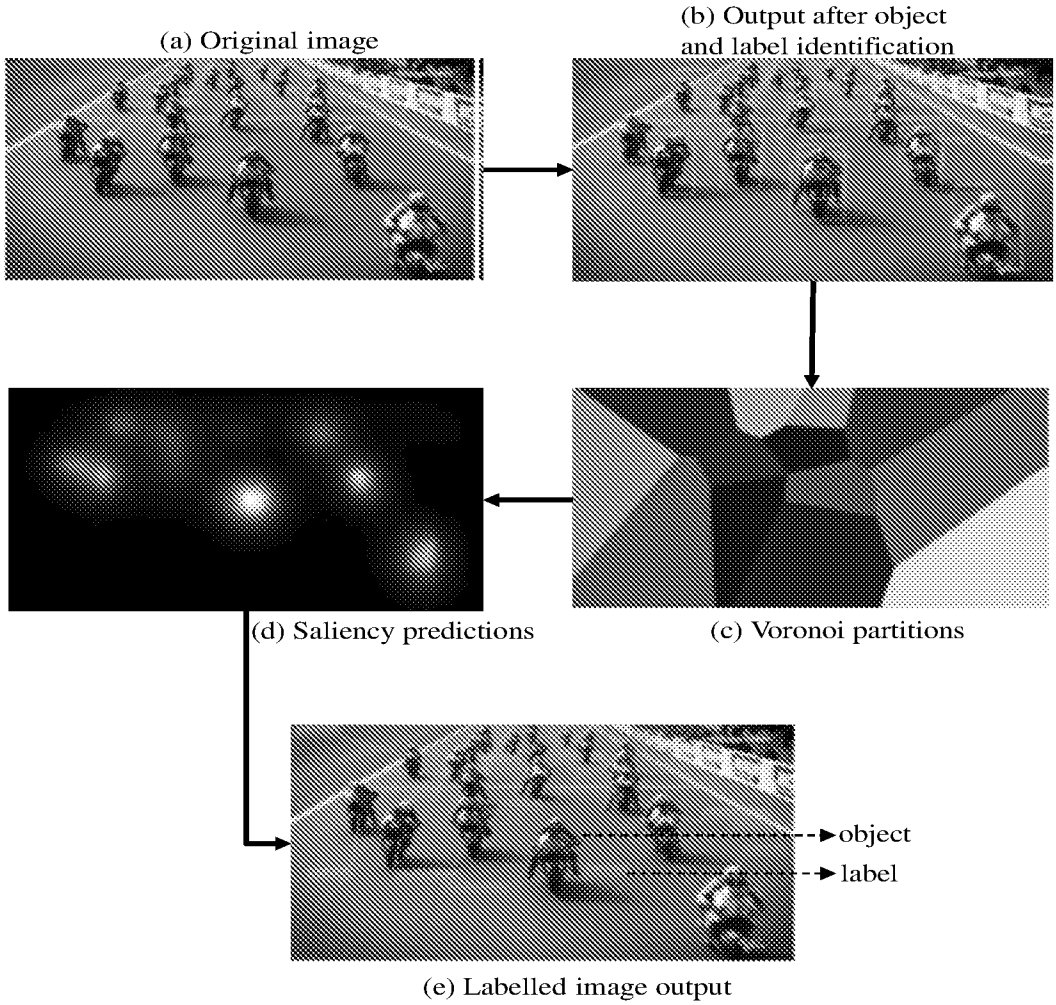

(a) Original image
(b) Output after object and label identification
(c) Voronoi partitions
(d) Saliency predictions
object
label
(e) Labelled image output

Fig. 5

INTERNATIONAL SEARCH REPORT

International application No.
PCT/IB 20/51706

A. CLASSIFICATION OF SUBJECT MATTER

IPC - G06T 11/00, G06T 7/00 (2020.01)

CPC - G06T 11/00, G06T 19/00, G06T 2219/004, G06T 2200/00, G06T 19/006

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
See Search History document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
See Search History document

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
See Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y --- A	US 7,131,060 B1 (Azum) 31 October 2006 (31.10.2006) entire document (especially col. 1, lines 61-63, col. 3, lines56-65, col. 6, lines 10-14, col. 6, lines 55-56, col. 8 lines 55-col. 9, line 11, col. 12, line 54-co.13, line 5).	1, 2, 5, 6, 9, 10 ---------- 3, 4, 7, 8, 11, 12
Y	US 2014/0359656 A1 (Adobe Systems Incorporated) 04 December 2014 (04.12.2014) entire document (especially Abstract & para [0008], [0025], [0026], [0075]-[0076], [0080], [0114], [0118], [0134] & claim 1).	1, 2, 5, 6, 9, 10 ---------- 3, 4, 7, 8, 11, 12
A	US 2012/0311496 A1 (Cao et al.) 06 December 2012 (06.12.2012) entire document (especially Abstract & para [0002], [0069], [0072], [0082])	3, 4, 7, 8, 11, 12
A	US 2008/0123945 A1 (Andrew et al.) 29 May 2008 (29.05.2008) entire document (especially para [0074]-[0077], [0338]).	3, 4, 7, 8, 11, 12
A	US 2003/0234782 A1 (Terentyev et al.) 25 December 2003 (25.12.2003) entire document (especially para	1-12

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"D" document cited by the applicant in the international application
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
21 June 2020 (21.06.2020)	15 JUL 2020
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	Authorized officer Lee Young Telephone No. PCT Helpdesk: 571-272-4300