

(51) International Patent Classification:
G06T 7/00 (2006.01)(21) International Application Number:
PCT/EP2015/066540(22) International Filing Date:
20 July 2015 (20.07.2015)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
14306224.8 31 July 2014 (31.07.2014) EP(71) Applicant: THOMSON LICENSING [FR/FR]; 1-5 rue
Jeanne d'Arc, 92130 Issy-lès-Moulineaux (FR).(72) Inventors: JACHALSKY, Joern; Landwehr 8, 30974
Wennigsen (DE). SCHEWZOW, Andrej; Talaverastr. 14,
30163 Hannover (DE).(74) Agent: SCHMIDT-UHLIG, Thomas; Deutsche Thomson
OHG, European Patent Operations, Karl-Wiechert-Allee
74, 30625 Hannover (DE).(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, DK, DM,
DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT,
HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR,
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, SM, ST, SV, SY, TH, TJ, TM, TN,
TR, TT, TZ, UA, UG, US, UZ, VC, VN, 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, LI, 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))

(54) Title: METHOD AND APPARATUS FOR INTERACTIVE VIDEO SEGMENTATION

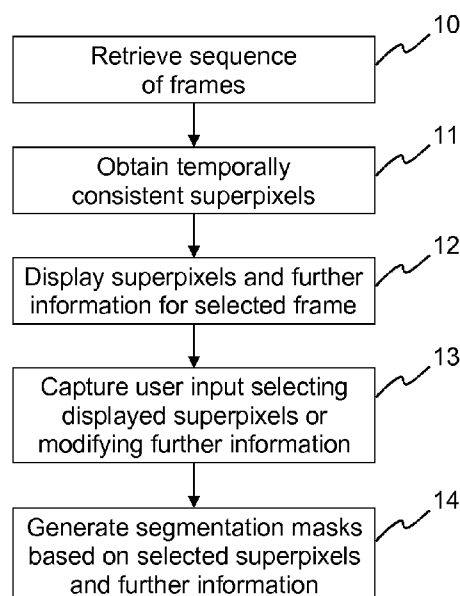


Fig. 20

(57) Abstract: A method and an apparatus for generating segmentation masks for a sequence of frames based on temporally consistent superpixels are described. A sequence of frames is retrieved (10) via an input (21). A superpixel unit (22) obtains (11) temporally consistent superpixels for the sequence of frames. Via a display unit (23) temporally consistent superpixels and further information related to the displayed superpixels for a selected frame from the sequence of frames are displayed (12) to a user. A user interface (24) captures (13) a user input selecting one or more of the displayed superpixels or modifying at least part of the further information related to the displayed superpixels. Using the selected one or more superpixels a segmentation mask generator (25) generates (14) segmentation masks for the sequence of frames.

METHOD AND APPARATUS FOR INTERACTIVE VIDEO SEGMENTATION**FIELD**

5 The present solution relates to a method and an apparatus for interactive video segmentation. More specifically, a method and an apparatus for generating segmentation masks for a sequence of frames based on temporally consistent superpixels are described.

BACKGROUND

10 Video segmentation is complex and often time- and memory-consuming, especially for high-resolution images. Superpixel algorithms represent a very useful and increasingly popular preprocessing step for video segmentation, but also for a wide range of other computer vision applications, such as tracking, multi-view object segmentation, scene flow, 3D layout estimation of indoor scenes, interactive scene modeling, image parsing, and semantic segmentation. Grouping similar pixels into so called superpixels leads to a major reduction of the image primitives. This results in an increased computational efficiency for subsequent processing steps, allows for more complex algorithms computationally infeasible on pixel level, and creates a spatial support for region-based features. Temporally consistent superpixels, as described in [1], help to reduce the complexity.

SUMMARY

30 It is an object of the present solution to provide an efficient tool for interactive video segmentation based on temporally consistent superpixels.

According to one embodiment, a method for generating segmentation masks for a sequence of frames based on temporally consistent superpixels comprises:

- retrieving a sequence of frames;
- 5 - obtaining temporally consistent superpixels for the sequence of frames;
- displaying temporally consistent superpixels and further information related to the displayed superpixels for a selected frame from the sequence of frames to a user;
- 10 - capturing a user input selecting one or more of the displayed superpixels or modifying at least part of the further information related to the selected superpixels; and
- generating segmentation masks for the sequence of frames using the selected one or more superpixels and the further
- 15 information related to the selected superpixels.

Accordingly, a computer readable storage medium has stored therein instructions enabling generating segmentation masks for a sequence of frames based on temporally consistent

20 superpixels, which when executed by a computer, cause the computer to:

- retrieve a sequence of frames;
- obtain temporally consistent superpixels for the sequence of frames;
- 25 - display temporally consistent superpixels and further information related to the displayed superpixels for a selected frame from the sequence of frames to a user;
- capture a user input selecting one or more of the displayed superpixels or modifying at least part of the further
- 30 information related to the selected superpixels; and
- generate segmentation masks for the sequence of frames using the selected one or more superpixels and the further information related to the selected superpixels.

Also, in one embodiment an apparatus configured to generate segmentation masks for a sequence of frames based on temporally consistent superpixels comprises:

- an input configured to retrieve a sequence of frames;
- 5 - a superpixel unit configured to obtain temporally consistent superpixels for the sequence of frames;
- a display unit configured to display temporally consistent superpixels and further information related to the displayed superpixels for a selected frame from the sequence of frames to
- 10 a user;
- a user interface configured to capture a user input selecting one or more of the displayed superpixels or modifying at least part of the further information related to the selected superpixels; and
- 15 - a segmentation mask generator configured to generate segmentation masks for the sequence of frames using the selected one or more superpixels and the further information related to the selected superpixels.

20 In another embodiment, an apparatus configured to generate segmentation masks for a sequence of frames based on temporally consistent superpixels comprises a processing device and a memory device having stored therein instructions, which, when executed by the processing device, cause the apparatus to:

- 25 - retrieve a sequence of frames;
- obtain temporally consistent superpixels for the sequence of frames;
- display temporally consistent superpixels and further information related to the displayed superpixels for a selected
- 30 frame from the sequence of frames to a user;
- capture a user input selecting one or more of the displayed superpixels or modifying at least part of the further information related to the selected superpixels; and

- generate segmentation masks for the sequence of frames using the selected one or more superpixels and the further information related to the selected superpixels.

- 5 Preferably, information on selected superpixels is provided in a superpixel table. This table gives an easily accessible overview on the selected superpixels to the user and can be used to manipulate the selection of superpixels.
- 10 The proposed solution introduces a fast way to interactively segment video sequences and generate segmentation masks. The selection and tracking of regions in frame sequences is based on temporally consistent superpixels, which are obtained, for example, by applying a superpixel algorithm to the sequence of
- 15 frames or by retrieving existing temporally consistent superpixels provided for the sequence of frames. The region selection using the displayed superpixels is very intuitive and easy to handle by the user. The video segmentation process can be split into two steps, i.e. an automatic offline-processing
- 20 (batch-processing) for superpixel generation and a real-time interactive video segmentation using these superpixels.

Segmentation masks for frames of the sequence of frames other than the selected frame are generated using label identifiers

25 of the selected superpixels. In this way the temporal consistency of the superpixels is used to propagate the selected regions across the subsequent frames of the sequence.

In one embodiment, one or more start frames and end frames of

30 the sequence of frames are set for a superpixel to limit tracking of the superpixel to selected ranges of frames. This allows the user to restrict tracking to a subsequence of the sequence of frames. In this way the user may accurately specify

which superpixel shall be considered at which point in time for generating a segmentation mask.

In one embodiment, user inputs to select a further superpixel for a frame of the sequence of frames other than the selected frame or to remove a selected superpixel are captured. Each further selected superpixel is added to the superpixel table with the start frame set to the current frame. Thereby, the solution allows the user to interactively refine the tracked/propagated regions on frame level. Removing a superpixel will completely remove a superpixel from tracking.

In one embodiment, user inputs to group two or more of the selected superpixels are captured. By grouping selected superpixels it becomes possible to distinguish different regions during the generation of the segmentation masks.

Preferably, information on selected superpixels is stored in a file. This information can be used as input for subsequent processing steps and allows resuming the superpixel selection at a later time. Alternatively or in addition, the generated segmentation masks are made available via an output or stored, e.g. as image files. Also these segmentation masks can be used as input for subsequent processing steps.

For a better understanding the present solution shall now be explained in more detail in the following description with reference to the figures. It is understood that the solution is not limited to this exemplary embodiment and that specified features can also expediently be combined and/or modified without departing from the scope of the present solution as defined in the appended claims.

BRIEF DESCRIPTION OF THE DRAWINGS

- Fig. 1 shows an example of a frame;
- 5 Fig. 2 depicts a superpixel label map corresponding to the frame of Fig. 1;
- Fig. 3 depicts the main elements of a graphical user interface of a video segmentation tool
- 10 Fig. 4 illustrates the GUI showing the first frame of a sequence with the superpixel boundaries as overlay;
- 15 Fig. 5 depicts the GUI of Fig. 4 toggled to an original view;
- Fig. 6 shows navigation and zoom buttons of the GUI;
- 20 Fig. 7 shows a superpixel table with exemplary superpixels;
- Fig. 8 depicts highlighted selected superpixels in a frame, whose end frame number is identical to the current frame number;
- 25 Fig. 9 illustrates grouping of superpixels;
- Fig. 10 shows grouped superpixels in the superpixel table;
- 30 Fig. 11 depicts a segmentation mask resulting from two selected groups of superpixels;

Figs. 12 to 16 illustrate the selection of an object in a frame;

Fig. 17 shows a selected region after setting an end frame;

5

Fig. 18 depicts a group resulting from grouping the superpixels of the selected region of Fig. 17;

Fig. 19 shows exemplary segmentation masks obtained for the group of Fig. 18;

10

Fig. 20 schematically illustrates one embodiment of a method a for video segmentation;

15

Fig. 21 shows a first embodiment of an apparatus configured to implement the method of Fig. 20; and

Fig. 22 schematically illustrates a second embodiment of an apparatus configured to perform the method of Fig. 20.

20

DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

25 In the following an exemplary implementation of the proposed solution shall be described. The implementation is an interactive video segmentation tool programmed in Python with a Qt-graphical user interface (GUI). The tool is both suited for computers with a mouse and a keyboard as well as for tablet
30 computers with touchscreens using touch gestures instead of mouse clicks.

The implemented solution requires as input a frame sequence and a corresponding sequence of superpixel label maps. These
35 superpixel label maps can be generated using, for example, the

algorithm described in [1], either beforehand in an independent superpixel generating step or upon reception of the frame sequence by the interactive video segmentation tool. Fig. 1 shows an example of a frame, whereas Fig. 2 depicts a corresponding superpixel label map. The superpixel labels are coded by grey values.

Fig. 3 depicts the main elements of the GUI 1 of the tool. The largest part of the GUI 1 is occupied by a frame area 2.

Located above the frame area 2 is a button area 3 comprising a variety of buttons. On the right side of the frame area 2 there is a superpixel table 4, which shows information about selected superpixels.

After loading the frame sequence and the corresponding superpixel maps into the tool the frame area 2 shows the first frame of the sequence with the superpixel boundaries as overlay. This is illustrated in Fig. 4. It is possible to toggle between the overlay view and an original view depicted in Fig. 5 by pressing a specific key on a keyboard or clicking a button in the GUI 1.

As can be seen in Fig. 6, the tool allows to playback, pause or go step by step through the sequence by clicking the appropriate buttons 5 or using keyboard shortcuts. For the navigation through the sequence also a slider 6 that is below the navigation buttons 1 can be used. Furthermore, with the zoom buttons 7 it is possible to zoom in and out, bring the view to the original size of the frame again or fit it to the current window size.

After navigating to the right frame in the sequence the user can start with the interactive video segmentation. To segment an object the user just has to select the region of the object.

The selection of a frame region is based on the selection of superpixels. There are two ways to select superpixels. The first one is to left-click (click with the left mouse button) on a superpixel. Selected superpixels will be highlighted with the color white and added to a superpixel table on the right side of the tool. The second way is helpful for continuous selecting. Dragging the mouse with the left mouse button clicked over the superpixels continuously selects them. The selected superpixels will be highlighted with the color white and added to the superpixel table 4 on the right side after release of the mouse button.

If a wrong superpixel has been selected, it can be deselected. Deselecting superpixels works in a similar way as selecting them. The only difference is that it is additionally necessary to press the shift key and then left-click the superpixel to remove it from the selection. It also works for continuously deselecting superpixels in case the shift key is held pressed and the mouse is dragged holding the left mouse button down over the superpixels that should be deselected. Deselected superpixels will also be removed from the superpixel table 4 on the right side.

For a precise selection the user can zoom into the frame or toggle between the original view and the overlay view. For each selected superpixel, the group identifier, the label identifier and the start as well as the end frame are indicated in the superpixel table 4.

Fig. 7 shows the superpixel table 4 with exemplary superpixels in more detail. It contains the following information about the selected superpixels:

- group number;

- label identifier of the superpixel;
- start and end frame numbers.

The label of the superpixel is an identifier for the temporally
5 consistent superpixel. It is calculated, for example, using the
unique RGB color of the superpixel in the superpixel label map.

The start and end frame number for a superpixel indicate the
(sub)sequence of frames, i.e. the time slot, for which the
10 superpixel should be tracked. When selecting a superpixel, it
is automatically added to the superpixel table 4 and its start
and end frame numbers are set in the following way:

- start frame number is set to the current frame number; and
- 15 - end frame number is set to the frame number of the last frame
in the sequence.

To change the start frame number of a selected superpixel the
user can simply navigate to a different frame in the sequence
20 and left-click the superpixel. The new start frame number will
be set. By holding the left mouse button down and dragging the
mouse over selected superpixels the user can change the start
frame numbers of multiple superpixels at once.

25 Changing the end frame numbers works in the same way as
changing the start frame numbers, the only difference is that
the user has to right-click the superpixel.

It is likewise possible to edit the start and end frame numbers
30 directly in the superpixel table.

As a support for the user, selected superpixels in the frame,
whose end frame number is identical to the current frame
number, are highlighted using the unique label grey value of

the superpixels. An example is depicted in Fig. 8, where the highlighted superpixels are those in the hat of the mannequin visible in the area identified by the white rectangle.

5 The label identifier of the superpixels is used to propagate the selected region across subsequent frames of the frame sequence. Thus in subsequent frames the superpixels with the same identifier are also selected. Stepping forward, using play, or the slider to navigate to a subsequent frame shows the
10 propagation of the selected region. The start and end frame can be used to refine the selection in the subsequent frames. Setting the end frame for a superpixel to frame number k excludes it from tracking for the frames with frame number $k+1$ and higher. Moreover, it is possible to add new superpixels in
15 subsequent frames. This is done in the same way as the initial selection. With these two methods the user has the full control to refine the propagated region. A superpixel can have multiple time slots, each time slot having its own start and end frame number. Thus, it is not only possible to exclude a superpixel
20 from the tracking at frame $j+1$. It is also possible to reinclude it in the tracking at frame $j+1+l$ with $l>0$. This is especially advantageous, for example, for superpixels that erroneously happen to switch from one object to another one and back. Using multiple time slots these tracking errors can be
25 handled.

The video segmentation tool is not restricted to handle only one region. Different regions can be identified using the group number. The group number preferably is an integer value between
30 1 and 255. It is used, for example, during the generation of the segmentation masks to distinguish the different regions. By default, the group number is 1. If the user does not change the group number, all selected superpixels will have a grey value of 1 in the generated segmentation masks. If the user wants to

create segmentation masks with multiple separate regions, the group feature of the tool should be used. Figs. 9 to 11 show an example in which the hats of the two mannequins on the right are tracked and each region gets its own group number. Figs. 9 and 10 show the process of setting the group identifier for the hat of the mannequin in the middle. To create a group the user selects appropriate superpixels in the superpixel table 4 and clicks the 'Group' button below the superpixel table 4. As a visual help the superpixels selected in the table are highlighted in light grey in the view, as visible in the area identified by the white rectangle. In a group dialog that appears when the 'Group' button is selected the user enters an integer value between 1 and 255. This group identifier can be used, for example, as a grey value in the segmentation mask. As visible in Fig. 10, subsequently the grouped superpixels are identified by their associated group number in the superpixel table 4. Fig. 11 depicts the segmentation mask for the displayed frame resulting from the two selected groups.

In order to remove all previously selected superpixels from the superpixel table 4 the user has the possibility to click a 'Reset' button below the superpixel table 4.

The segmentation tool provides the functionality to assess the propagation of the selected regions. Thereby, the navigation features (play, pause, step, and slider) play a central role. The user can (re-)play the complete sequence and pause at frames, in which the tracked regions need a further inspection, or simply step directly through the complete sequence. For the inspection, the zoom-feature as well as the switching of the views is helpful.

After reviewing and potentially refining the tracked regions, the user can export them as either text files or segmentation

masks, which are generated as grey-scale images. The generated text file will contain information about the selected superpixels. For example, for each selected superpixel a new line with group number, label identifier, and start and end frame numbers for each time slot is added. A text file with one selected superpixel in two time slots would thus look as follows:

#	group	label	start1	end1	start2	end2
10	100	77136081	2	9	14	19

The regions exported as text files can be loaded into the tool again. This is especially useful if the user wants to resume the segmentation work at a later point in time, share the work with others, or create multiple differing versions.

For exporting the selected regions, i.e. the selected superpixels, as a sequence of segmentation masks, the user has to click the 'Image' button below the superpixel table 4. In a dialogue that opens, the user then chooses an output directory, a bit depth for the grey-scale images (8 bit or 24 bit) and then clicks 'Start'. After successful completion of the processing the generated images are available in the output directory.

25

In the following a brief workflow example shall be described. A short sequence with 20 original frames and the corresponding superpixel label maps are used. In this example, the selection of the region should begin in the third frame and the tracking should stop in frame 17. After loading the project the view depicted in Fig. 4 appears. Using the navigation buttons (or the slider) the user navigates to the third frame. The user then selects the superpixels covering the dress of the mannequin in the middle. The selection process is depicted in

30

Figs. 12 to 16. The selected superpixels are automatically added to the superpixel table. As the internal frame numbers start with 0, their start frame number is automatically set to 2 and their end frame number to the end of sequence, which in this example is 19.

Based on the superpixel labels, this selection is now propagated across the subsequent frames until the end of the sequence. In order to control whether the selection is correctly propagated, i.e. if the right superpixels are also selected in the subsequent frames, the user can navigate through the sequence. Thereby, it is possible to refine the selection as described further above.

After the inspection the end frame number should be set as intended to stop the tracking of the selected superpixels in the frame 17. The end frame number is either set by right-clicking the superpixels in frame 17 or by directly editing their end frame number in the superpixel table. For a visual help, the superpixels, whose end frame number is equal to the frame number of the displayed frame, are highlighted using the unique label grey value. After setting the end frame to frame 17 the selected region looks as shown in Fig. 17.

Once the selected region is correct over the frames, groups are created and unique numbers are set for different selected regions. In the present example, however, there is only one region. The user selects the lines with the superpixels belonging to a region in the superpixel table 4 and clicks the 'Group' button. In the present case these are all lines in the table. The user then enters a group number and clicks OK. The resulting group is depicted in Fig. 18.

When each region has a unique group number, the segmentation masks can be exported as described above. In the present case, the generated segmentation masks look like illustrated in Fig. 19. In this figure not all segmentation masks are shown.

5

One embodiment of a method for generating segmentation masks for a sequence of frames based on temporally consistent superpixels is schematically illustrated in Fig. 20. In a first step a sequence of frames is retrieved 10, e.g. from a network or from a local storage. Temporally consistent superpixels for the sequence of frames are then obtained 11, e.g. by applying a superpixel algorithm to the sequence of frames or by retrieving existing temporally consistent superpixels provided for the sequence of frames. Once the temporally consistent superpixels for the sequence of frames are available, the superpixels for a selected frame and further information related to the displayed superpixels are displayed 12 to a user. The method proceeds with capturing 13 a user input selecting one or more of the displayed superpixels or modifying at least part of the further information related to the selected superpixels. Finally, using the selected one or more superpixels and the further information related to the selected superpixels segmentation masks are generated 14 for the sequence of frames.

Fig. 21 schematically illustrates one embodiment of an apparatus 20 for generating segmentation masks for a sequence of frames based on temporally consistent superpixels. The apparatus 20 comprises an input 21 for retrieving 10 a sequence of frames, e.g. from a network or from a local storage. A superpixel unit 22 obtains 11 temporally consistent superpixels for the sequence of frames, e.g. by applying a superpixel algorithm to the sequence of frames or by retrieving existing temporally consistent superpixels provided for the sequence of frames. Via a display unit 23, e.g. a display device or an

output connected to a display device, temporally consistent superpixels and further information related to the displayed superpixels for a selected frame from the sequence of frames are displayed 12 to a user. The apparatus further comprises a user interface 24 for capturing 13 a user input selecting one or more of the displayed superpixels or modifying at least part of the further information related to the selected superpixels. Using the selected one or more superpixels and the further information related to the selected superpixels a segmentation mask generator 25 generates 14 segmentation masks for the sequence of frames. The resulting segmentation masks are preferably stored on a local storage 26 or made available at an output 27. The superpixel unit 22, the segmentation mask generator 25, and the user interface 24 may likewise be fully or partially combined into a single unit or implemented as software running on a processor. In addition, the user interface 24 may be part of the display unit 23, e.g. in the form of a touch screen. Also, the input 21 and the output 27 can likewise form a single bi-directional interface.

Another embodiment of an apparatus 30 configured to perform the method for generating segmentation masks for a sequence of frames based on temporally consistent superpixels is schematically illustrated in Fig. 22. The apparatus 30 comprises a processing device 31 and a memory device 32 storing instructions that, when executed, cause the apparatus to perform steps according to one of the described methods.

For example, the processing device 31 can be a processor adapted to perform the steps according to one of the described methods. In an embodiment said adaptation comprises that the processor is configured, e.g. programmed, to perform steps according to one of the described methods.

A processor as used herein may include one or more processing units, such as microprocessors, digital signal processors, or combination thereof.

- 5 The local storage and the memory device 32 may include volatile and/or non-volatile memory regions and storage devices such hard disk drives and DVD drives. A part of the memory is a non-transitory program storage device readable by the processing device 31, tangibly embodying a program of instructions
- 10 executable by the processing device 31 to perform program steps as described herein according to the present principles.

References

- [1] M. Reso et al.: "*Temporally Consistent Superpixels*",
International Conference on Computer Vision (ICCV), 2013,
pp. 385-392.

CLAIMS

1. A method for generating segmentation masks for a sequence of frames based on temporally consistent superpixels, the
5 method **comprising**:
 - retrieving (10) a sequence of frames;
 - obtaining (11) temporally consistent superpixels for the sequence of frames;
 - displaying (12) temporally consistent superpixels and
10 further information related to the displayed superpixels for a selected frame from the sequence of frames to a user;
 - capturing (13) a user input selecting one or more of the displayed superpixels or modifying at least part of the further information related to the selected superpixels; and
15 - generating (14) segmentation masks for the sequence of frames using the selected one or more superpixels and the further information related to the selected superpixels.
2. The method according to claim 1, **further** comprising
20 providing information on selected superpixels in a superpixel table (4).
3. The method according to claim 1 or 2, **wherein** segmentation masks for frames of the sequence of frames other than the
25 selected frame are generated using label identifiers of the selected superpixels.
4. The method according to one of claims 1 to 3, **further**
30 comprising setting one or more start frames and end frames in the sequence of frames for a superpixel to limit tracking of the superpixel to selected ranges of frames.
5. The method according to one of the preceding claims, **further** comprising capturing a user input to select a further

superpixel for a frame of the sequence of frames other than the selected frame or to remove a selected superpixel.

6. The method according to one of the preceding claims, **wherein**
the temporally consistent superpixels for the sequence of
frames are retrieved (11) by applying a superpixel algorithm
to the sequence of frames or by retrieving existing
temporally consistent superpixels provided for the sequence
of frames.

7. The method according to one of the preceding claims, **further**
comprising capturing a user input to group two or more of
the selected superpixels.

8. The method according to one of the preceding claims, **further**
comprising storing information on selected superpixels in a
file and/or storing the generated segmentation masks as
image files.

9. A computer readable storage medium having stored therein
instructions enabling generating segmentation masks for a
sequence of frames based on temporally consistent
superpixels, which when executed by a computer, cause the
computer to:

- retrieve (10) a sequence of frames;
- obtain (11) temporally consistent superpixels for the
sequence of frames;
- display (12) temporally consistent superpixels and further
information related to the displayed superpixels for a
selected frame from the sequence of frames to a user;
- capture (13) a user input selecting one or more of the
displayed superpixels or modifying at least part of the
further information related to the selected superpixels; and
- generate (14) segmentation masks for the sequence of

frames using the selected one or more superpixels and the further information related to the selected superpixels.

10. An apparatus (20) configured to generate segmentation masks for a sequence of frames based on temporally consistent superpixels, **wherein** the apparatus (20) comprises:

- an input (21) configured to retrieve (10) a sequence of frames;

- a superpixel unit (22) configured to obtain (11)

temporally consistent superpixels for the sequence of frames;

- a display unit (23) configured to display (12) temporally consistent superpixels and further information related to the displayed superpixels for a selected frame from the sequence of frames to a user;

- a user interface (24) configured to capture (13) a user input selecting one or more of the displayed superpixels or modifying at least part of the further information related to the selected superpixels; and

- a segmentation mask generator (25) configured to generate (14) segmentation masks for the sequence of frames using the selected one or more superpixels and the further information related to the selected superpixels.

11. An apparatus (30) configured to generate segmentation masks for a sequence of frames based on temporally consistent superpixels, the apparatus (30) comprising a processing device (31) and a memory device (32) having stored therein instructions, which, when executed by the processing device (31), cause the apparatus (30) to:

- retrieve (10) a sequence of frames;

- obtain (11) temporally consistent superpixels for the sequence of frames;

- display (12) temporally consistent superpixels and further

information related to the displayed superpixels for a selected frame from the sequence of frames to a user;

- capture (13) a user input selecting one or more of the displayed superpixels or modifying at least part of the

5 further information related to the selected superpixels; and

- generate (14) segmentation masks for the sequence of frames using the selected one or more superpixels and the further information related to the selected superpixels.



Fig. 1

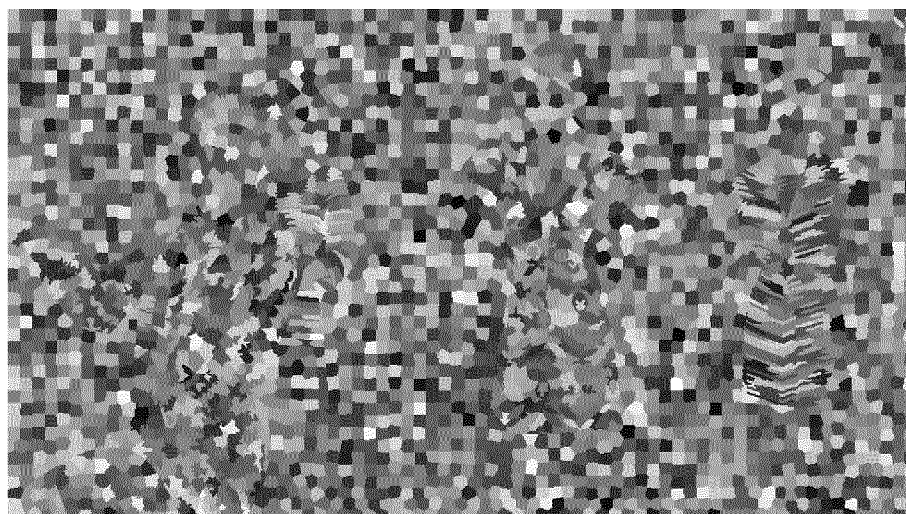


Fig. 2

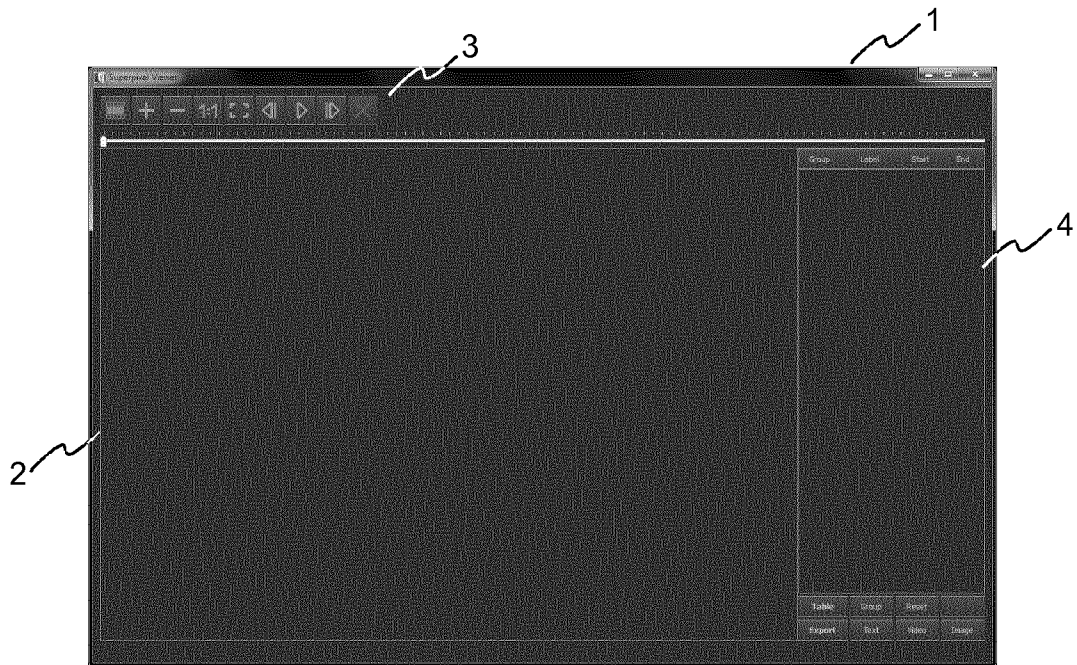


Fig. 3



Fig. 4

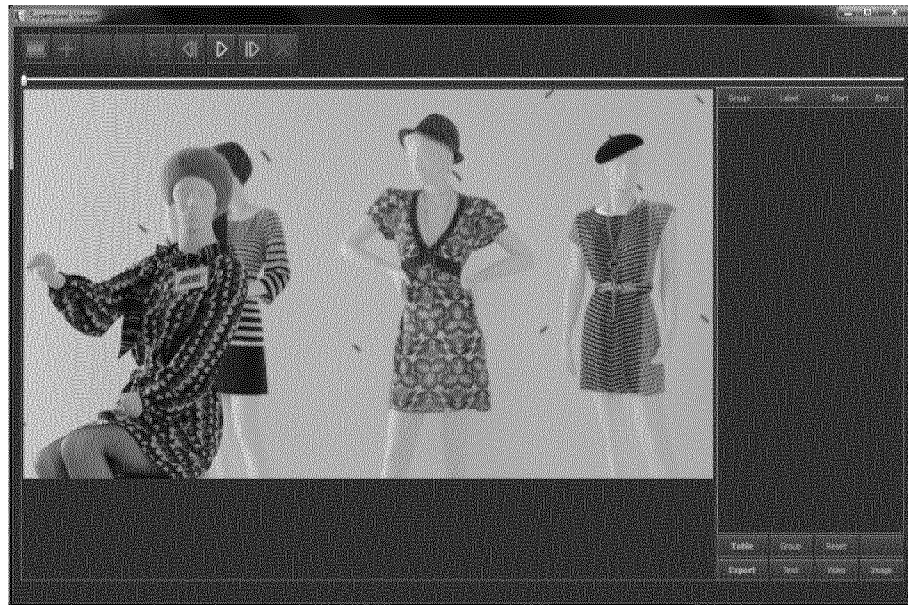


Fig. 5

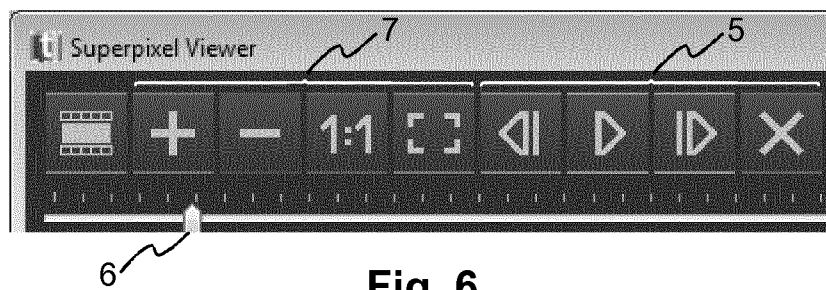


Fig. 6

Group	Label	Start	End
1	77136081	1	19
1	228121229	1	19
1	30250169	1	19
1	165156115	1	19
1	16129251	1	19
1	87106166	1	19

Table

Group

Reset

Export

Text

Video

Image

Fig. 7

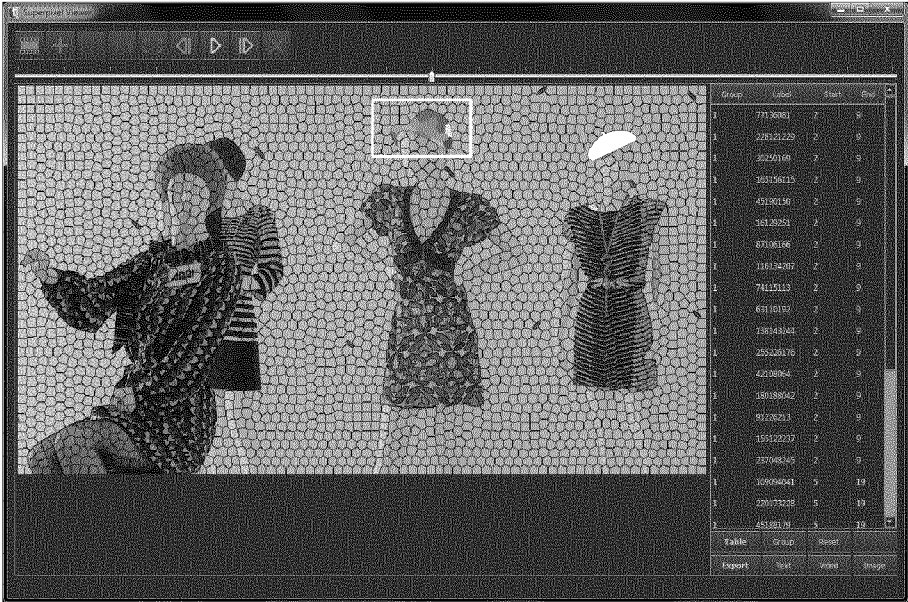


Fig. 8

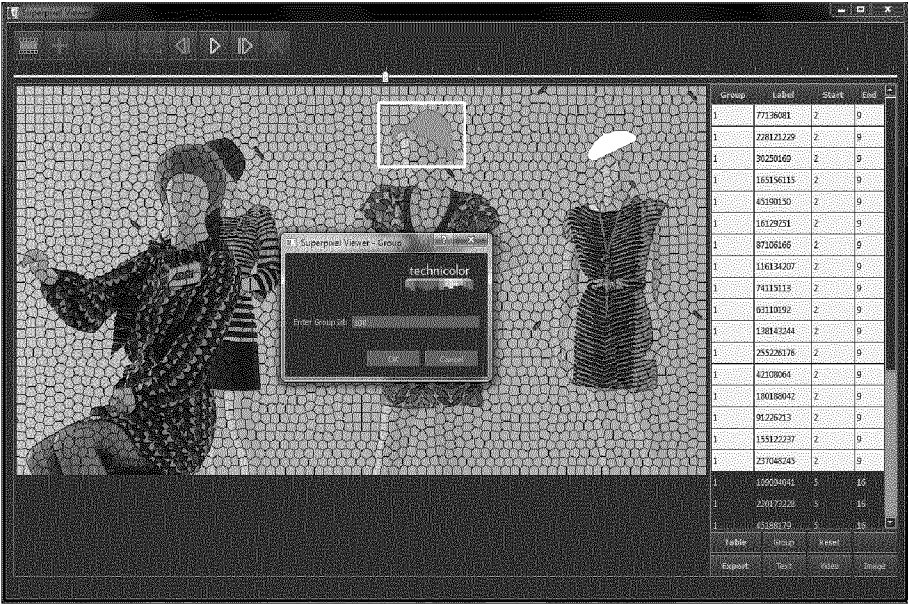


Fig. 9

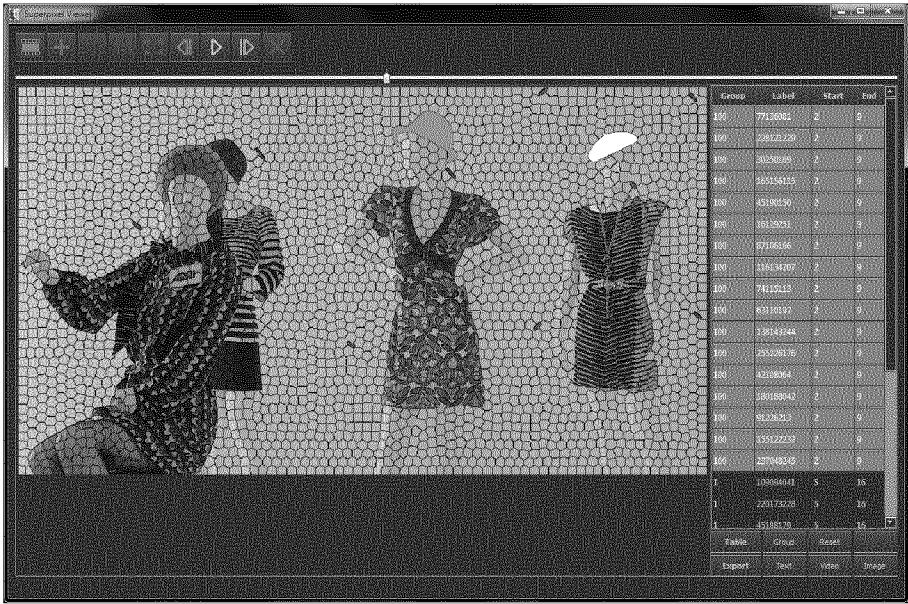
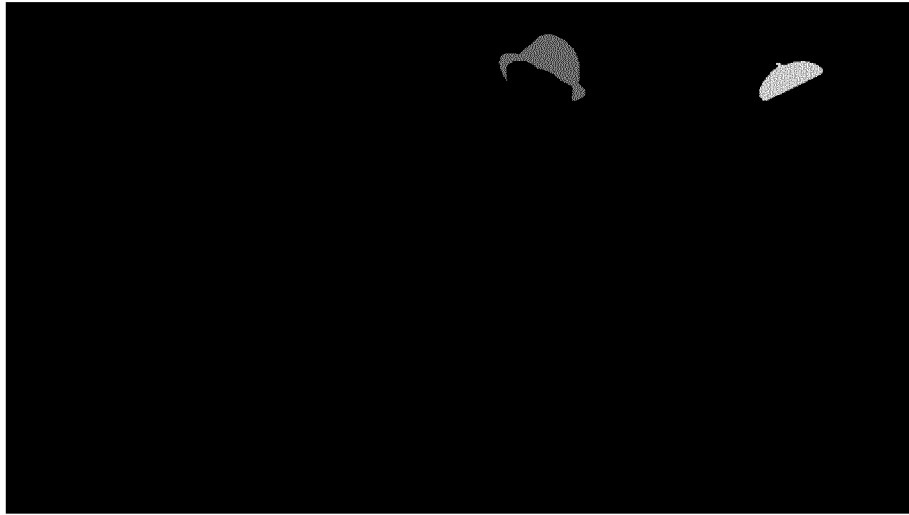


Fig. 10

**Fig. 11****Fig. 12**

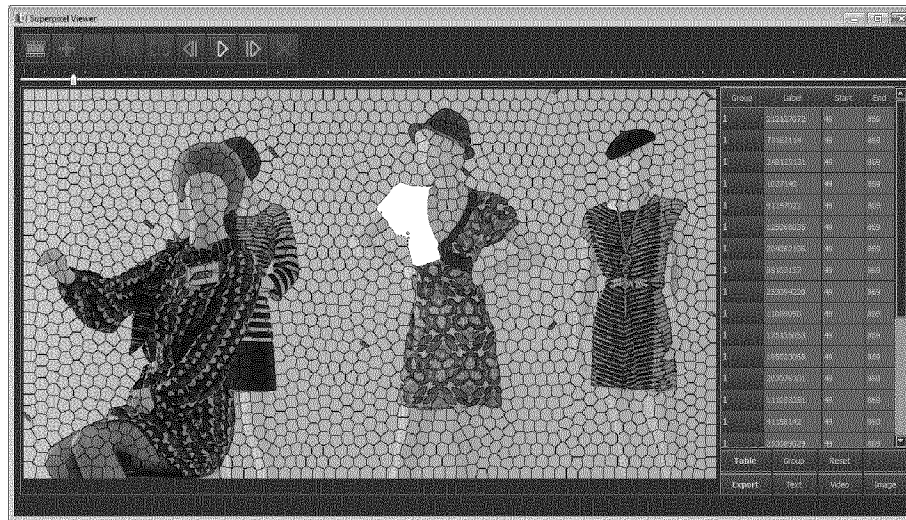


Fig. 13

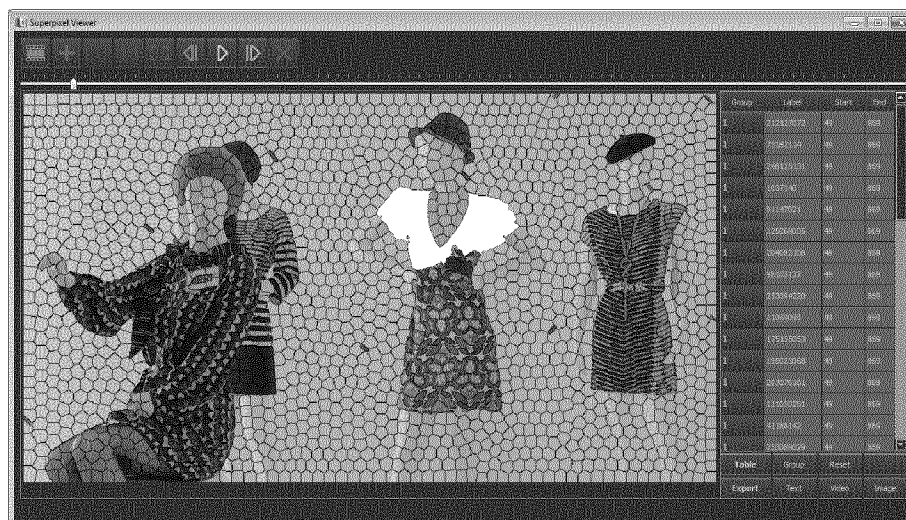


Fig. 14



Fig. 15



Fig. 16

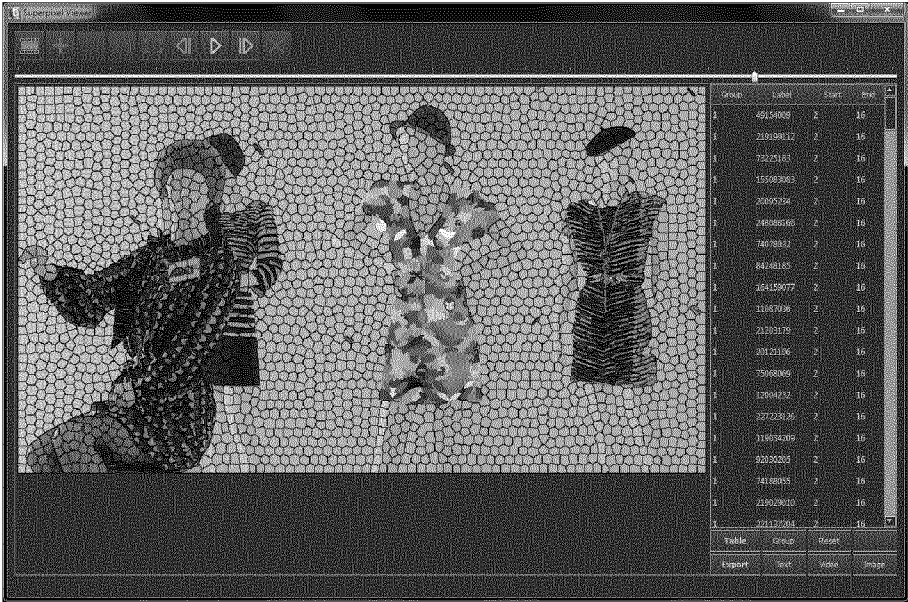


Fig. 17

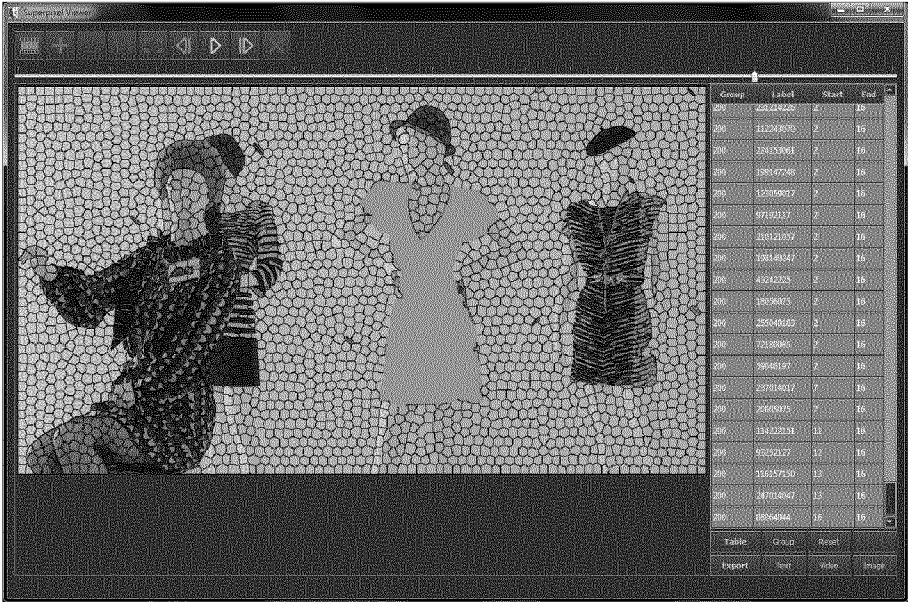
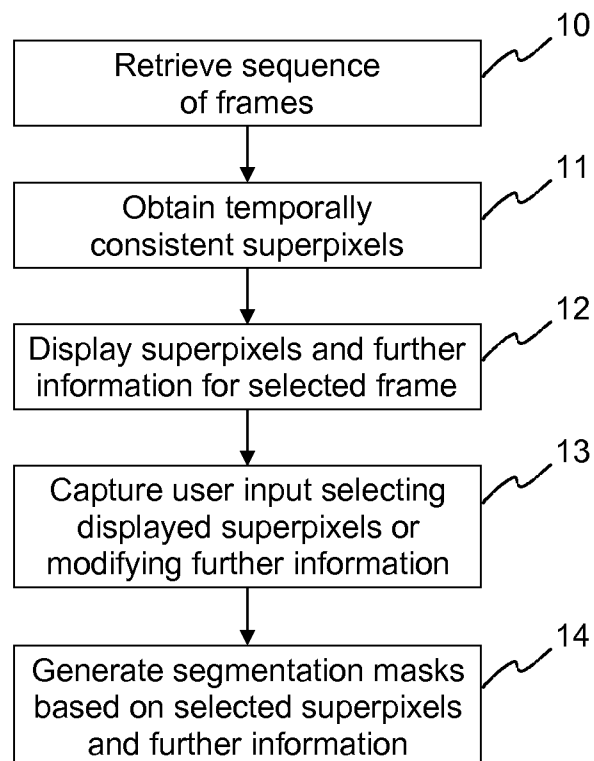
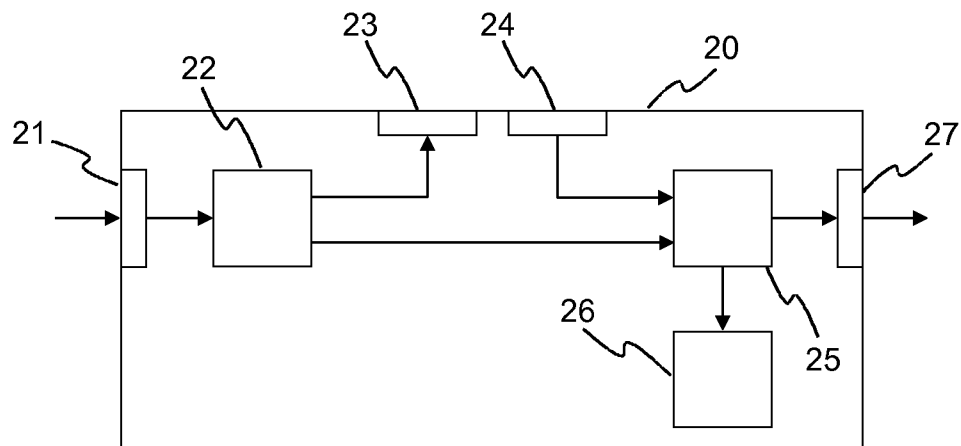
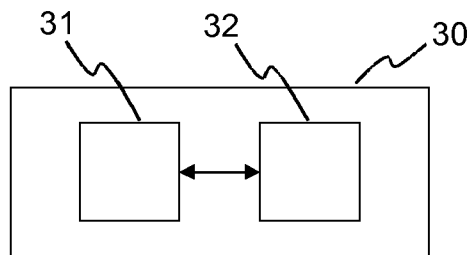


Fig. 18

**Fig. 19****Fig. 20**

**Fig. 21****Fig. 22**

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2015/066540

A. CLASSIFICATION OF SUBJECT MATTER
INV. G06T7/00
ADD.

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)
G06T G11B

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

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

EPO-Internal, COMPENDEX, INSPEC, IBM-TDB, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Johannes Furch ET AL: "D4.3.2 Hybrid Scene Analysis Algorithms", 30 October 2013 (2013-10-30), pages 1-60, XP055176149, Retrieved from the Internet: URL: http://3d-scene.eu/outcomes.htm [retrieved on 2015-03-12] section "The Superpixel Coherency Analyser (SCA)" figures 20-22 ----- -/--	1-3,5,6, 8-11



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

"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

17 September 2015

Date of mailing of the international search report

28/09/2015

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Engels, Angela

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2015/066540

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DONDERA RADU ET AL: "Interactive video segmentation using occlusion boundaries and temporally coherent superpixels", IEEE WINTER CONFERENCE ON APPLICATIONS OF COMPUTER VISION, IEEE, 24 March 2014 (2014-03-24), pages 784-791, XP032609947, DOI: 10.1109/WACV.2014.6836023 [retrieved on 2014-06-16] abstract section 2 figure 2	1,3-11
X	----- LIU Z ET AL: "Semi-automatic video object segmentation using seeded region merging and bidirectional projection", PATTERN RECOGNITION LETTERS, ELSEVIER, AMSTERDAM, NL, vol. 26, no. 5, 1 April 2005 (2005-04-01), pages 653-662, XP025292474, ISSN: 0167-8655, DOI: 10.1016/J.PATREC.2004.09.017 [retrieved on 2005-04-01] abstract figure 2	1,2,9-11
A	----- "Using Using Adobe Photoshop CS4 for Windows and Mac OS - Chapters 2 and 18" In: "Using Using Adobe Photoshop CS4 for Windows and Mac OS", 1 October 2010 (2010-10-01), Adobe Systems Incorporated, XP055212771, section "Work with the Info Panel" page 533, paragraph 2 -----	2