Labeling Tool - User Manual

Ricardo Ribeiro (ribeiro@isr.ist.utl.pt)

October 28, 2014

Contents

List of Figures

1   Software version

This manual documents the version 0.1.1 of the labeling tool (labelingtool v0.1.1).

2   Introduction

For the training and testing of vision algorithms it is necessary to use image sequences that where manually labeled.

The labeling task consists in tagging all objects in the image with a rectangle. Such operation needs to be done manually since the labeling is to be considered as the ground truth against which the results of the algorithms are compared. Any kind of error in the labeling process will have a negative impact in the performance of the algorithms and will compromise the reliability of the test results.

This manual process is quite heavy and long, specially if the number of images is large, as normally is the case for video sequences. For that reason, this tools was developed aiming to alleviate the weight of this task. This was achieved in several ways.

One way was to create an very light and fast interface that minimizes the number of operations necessary for each label. This was achieved by developing this tool in C++ (the OpenCV libraries where used). There was care into making the execution of all operations the fastest possible so that the user of the tool can immediately see the result of the commands executed, even if the commands are given in fast sequence. The manual tools that existed previously, developed in Matlab, where too slow.

Another way to alleviate the users work was to provide automatic mechanisms to create the labels without dispensing the user of the task of adjusting, at least, of verifying them. These automations try to place new label the closest possible of the intended final position for them. The user only needs to do small adjustments or to confirm the labels correctness.

This tool is generic and can be applied to label any kind of object in any kind of images. Its limitation is that only rectangular labels are supported.

3   Instalation and compilation of the tool

3.1   Obtaining the tool (SVN)

The tool is in the folder

```
devel/tools/labeling_tool
```

in the Seagull SVN repository located at the ISR servers. To download the tool use the command:

```
svn checkout svn://svn.isr.ist.utl.pt/seagull/devel/tools/labeling_tool
```

3.2   Dependencies and supported OS's

This tool depends on the OpenCV libraries which need to be installed before build time.

To build the tool is also necessary to have cmake installed.

The tool was developed in linux, so it compiles and works well in this operating system.

It was also compiled and tested in Mac OS X. This version is usable but has some problems in the window resizing features described in subsection 5.1.8.

3.3   Build the tool

To build the tool use the following commands:

```
cd labeling_tool
mkdir build
cd build
cmake ..
make
```

4   Tool execution

4.1   Command line

The tool is executed inside the build folder (see section 3.3) using the command

```
./labeling_tool [options] <video_file>
```

To obtain help use

```
./labeling_tool -h
```

4.2   Video file

The video file can be in any format supported by OpenCV (examples: .mpg, .avi, .mkv, etc.). Name the file in the command line. For example:

```
./labeling_tool lanchaArgos_clip1.avi
```

If the file is in a different folder, then also include the path.

```
./labeling_tool videos/lanchaArgos_clip1.avi
```

4.3   Labeling file

By default, the filename where the labels are saved is the same as the video file with its extension replace by .gt.txt. For example, the command

```
./labeling_tool videos/lanchaArgos_clip1.avi
```

creates the labeling file

```
videos/lanchaArgos_clip1.gt.txt
```

Alternatively, a different filename can be specified using the -d option.

```
./labeling_tool -d labels.gt.txt lanchaArgos_clip1.avi
```

The labeling file format (labelingtool v0.1) is a simple text file with one line per label. Each line has the following format:

```
<frame number> <x> <y> <width> <height> <object id> <1(temporary)/0(final)>
```

The file format is also documented in the tools's help.

```
./labeling_tool -h
```

4.4   Autosave e backup

Every 30 seconds a backup file with extension .autosave is saved with all labels and changes made so far. This file is deleted when the program exits normally.

There is also a backup file that is coppied from the original file before the program saves the new file. This backup file has the same filename as the original but with a tilde (˜) added (example: labels.gt.txt˜).

5   Using the tool

5.1   General use

5.1.1   Help ("h" key)

An help screen is always available at any moment (Figure 1). It contains all the key shortcuts available and a brief description of the corresponding commands. Press "h" to call the help screen and press "h" again to exit from it.

The tool continues to work normally even when the help screen is active. The screen itself is translucent and the mouse and keyboard are still active, so the user can keep working while seeing the help screen.

```
h — show/hide this help
v — show/hide frame number and current id info
x — show/hide box id's
1 — maximizes dynamic range of the image shown.
ESC — quits the program and saves detections to file

moving box borders (normal — move by 1 pixel; with shift — move by 10 pixels):
   3/e — top border up/down
   a/s — left border left/right
   f/g — right border left/right
   d/c — bottom border up/down

,/. — previous/next frame
m/— — previous/next 10th frame
;/: — previous/next 100th frame
M/_ — first/last frame
k/l — previous/next final box
K/L — previous/next estimated box
Space — start/stop play

5/6 — decrement/increment current id
+ — create a new box (at <10,10,100,100>)
* — delete current box
t — set current box as temporary
T — set current box as final
enter — set current box as final and advance to next frame

p — copy the box of previous frame an sets as temporary box (previous box needs to be final.)
o — search for the image within the previous frame box (needs to be final) in the current image and sets new temp box
i — interpolates between the previous final box and current box
u — adjust the temporary boxes between current and last final box using image search (both boxes, at start and at end, need to be final)

7/8/9 — resize window to half/full/double size
O — enter/exit full screem mode
Note: Final boxes of current id are green and temporary boxes are yellow. Boxes with other id's are grey.
```

Figure 1: Help - Press the "h" key.

5.1.2 Keyboard

The commands available in the tool are almost all executed through the use of the keyboard (see Figure 2). There are keys or combinations of keys to execute all operations. Ergonomics was taken into account when assigning the shortcut keys so that the number of necessary movements of the user hands and fingers was minimized, not only for single command but also for sequences of commands.

The keyboard map from Figure 2 is also available at

```
labeling_tool/doc/keyboard_map.pdf
```

in a file ready to be printed for reference.

Note: This tool was developed based on the Portuguese keyboard layout from Portugal. For other keyboard layouts the commands may be located differently.

5.1.3 Mouse

It is possible to use the mouse, but that option is more useful for large movements, which are rare. It is preferable to keep the hands on the keyboard instead of waisting time moving the hand between the mouse and the keyboard.

The mouse can be used to move or to modify a label by clicking and dragging. The point where the user clicks defines the operation. For more information see subsection 5.2.5.

5.1.4 Save

Contrary to what is usual, there is no specific command to save the labels file because the file is automatically saved when the program exits (when the user
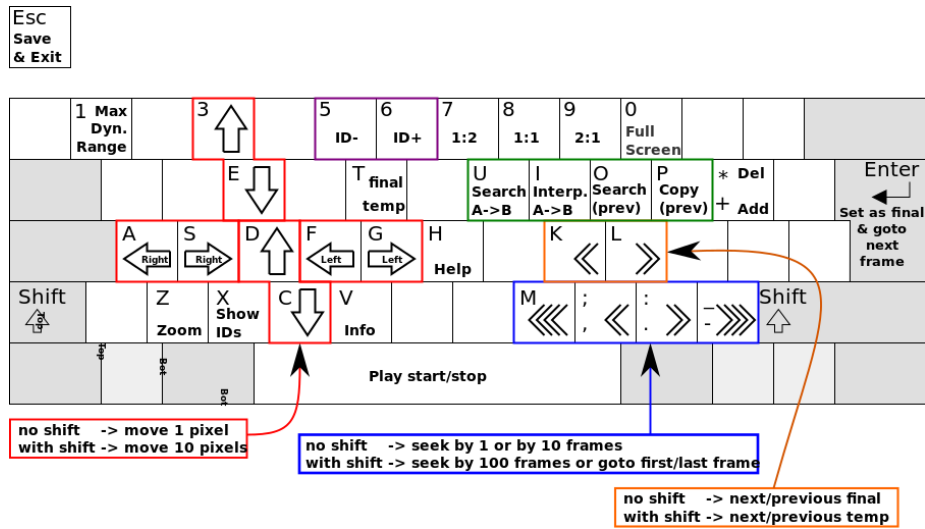
Figure 2: Keyboard map showing all available the commands (this map is available at labeling_tool/doc/keyboard_map.pdf).

preses the ESC key).[1]

Every 30 seconds, a file with .autosave extension is saved with all current changes. This file is deleted when the program exits normally.

5.1.5 Exit

Press the ESCAPE key. The labels file is saved automatically.

5.1.6 Navigate the video

There are many key to go forward and backward on the video file. The following is available:

- play and stop the video - SPACE key

- seek backward/forward by 1 frame - ",” and ".” keys

- seek backward/forward by 10 frames - "m” and "-” keys

- seek backward/forward by 100 frames - ";” and ":” keys (i.e., SHIFT + ",” and SHIFT + ".”)

- goto first frame - "M” key (i.e., SHIFT + "m”)

- goto last frame - "_” key (i.e., SHIFT + "-”)

Is is also possible to find the previous/next frame which:

- contains a temporary label - SHIFT + "k” and SHIFT + "l”

---

[1] **Tip:** to exit the program without saving go to the terminal where the program was started and press CTRL+”c”. In this way the program is immediately killed without having the opportunity to save the file.
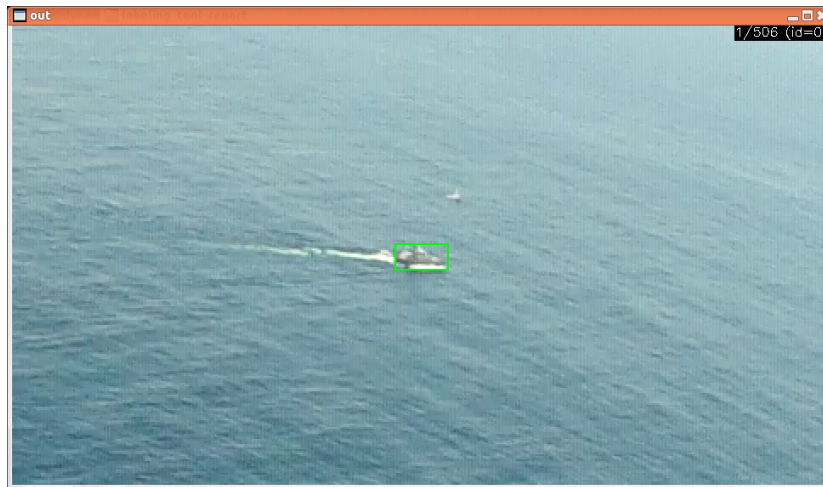
Figure 3: Information - by pressing the "v" key the frame number and object ID is shown in the upper right corner of the image.

- contains a final label - "k" and "l"

Press "v" to show the current frame number in the top right corner of the image (Figure 3).

5.1.7 Zoom

In order to achieve better accuracy, is useful to see in detail the area where the label is placed. Pressing the "x" key cyclically changes the zoom between 1 (no zoom), 2 and 4 times zoom (Figure 4). All commands work independently of the zoom selected.

The zoom is automatically centered around the marked label. Any adjustment to the label causes a new centering.

Note: This command does not change the window size. To do that see subsection 5.1.8.

Warning1: Due to the autocentering feature, the use of the mouse to change the label while zooming can be cumbersome.

Warning2: When seeking the video with the zoom enabled, the image can jump unpredictably. This behavior is normal and happens when a frame does not have a defined label, making the centering impossible in that frame. Solution: choose 1x zoom or ignore the image jumps.

5.1.8 Window size

(See subsection 3.2 about the limitation in Mac OS X.)

The keys "7", "8" e "9" allow to set the size of the window to half, full and double size compared to the original video size. This window size is independent of the zoom selected (see subsection 5.1.7).
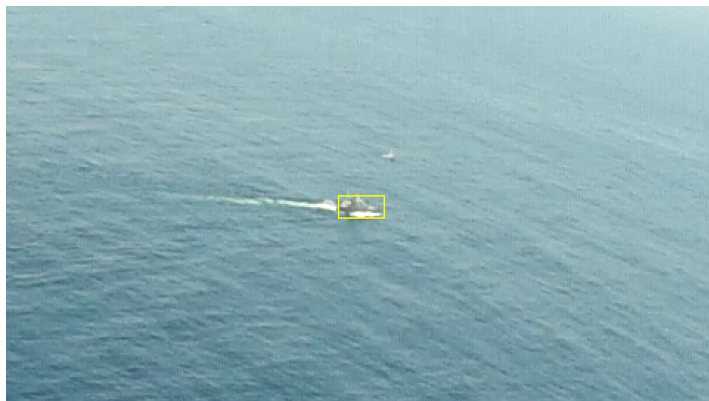
It is also possible to put the window in full screen by pressing the "0" key. In this mode the use of the screen is maximized.

(a) 1x

(b) 2x

(c) 4x

Figure 4: Zoom - press the "z" key to cyclically change the zoom: 1x→2x→4x(→1x→etc.).
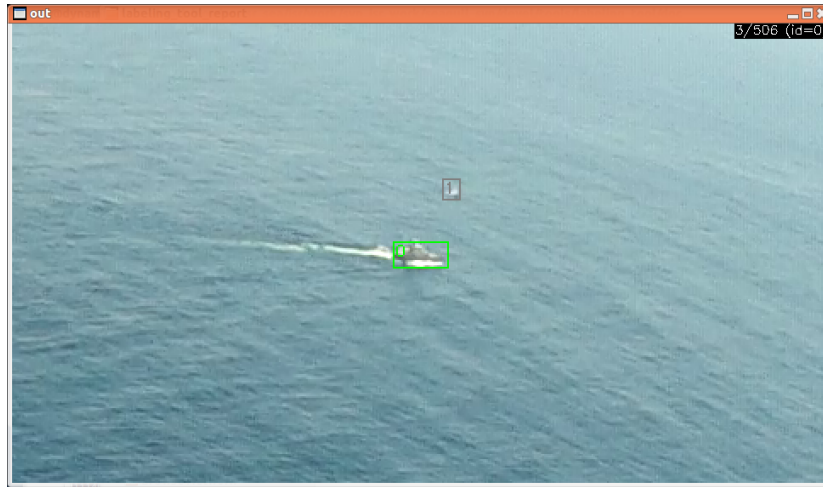
Figure 5: Object ID's - pressing "x" will show the ID's inside the label rectangles. Pressing "v" will show the current ID in the upper right corner (in this case, ID=0).

Note: To obtain a real pixel representation of the video on screen choose the original windows size ("8" key) and select 1x zoom (subsection 5.1.7).

Note: By default, the tools starts with half window size. This helps when the videos are larger than the screen size.

5.1.9 Object ID's

The tool is ready to label different objects on the same image. To each object is assigned a different identification number (ID). The current ID can be incremented or decremented using the "5" and "6" keys.

A label with an ID equal to the current ID has the color yellow or green according to the label type (see subsection 5.2.2). The labels with other ID's are shown in gray.

Pressing the "x" key, the ID's are all shown inside the labels they belong to (Figure 5 and Figure 6d). By pressing "v" the currently selected ID is shown in the top right corner of the image.

The commands only affect the label with the current ID. To create, modify or delete a label with a different ID it is mandatory to first use the "5" and "6" keys to select the correct ID.

5.2 Labels

5.2.1 Label anatomy

A label consists in a rectangle with four borders. The object should remain inside the label and the label should be the smallest possible that fully contains the object.

With this tool it is possible to move the label as a whole or to move its borders independently.
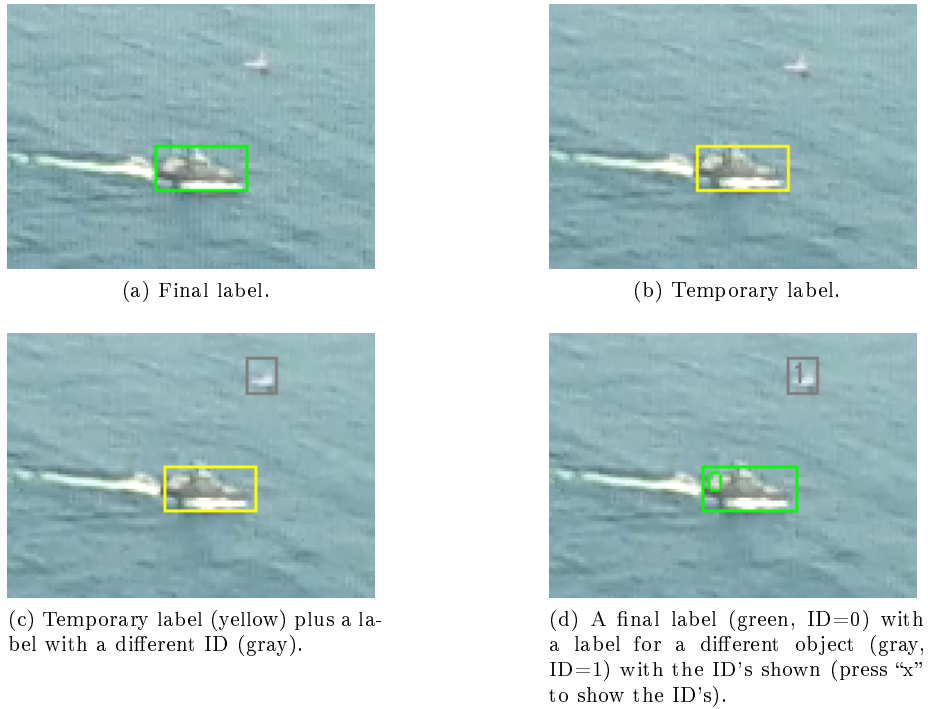
(a) Final label.

(b) Temporary label.

(c) Temporary label (yellow) plus a label with a different ID (gray).

(d) A final label (green, ID=0) with a label for a different object (gray, ID=1) with the ID's shown (press "x" to show the ID's).

Figure 6: Label types.

5.2.2   Label types

There are two types of labels (see Figure 6):

- temporary labels - yellow color

- final labels - green color

When a label is created it's type is by default temporary. This includes manually (subsection 5.2.3) or automatically (subsection 5.3) created labels.

The labels will only become final when:

- are adjusted by the user (subsection 5.2.5)

- are marked by the user as final (subsection 5.2.6)

Note: It was tried to minimize the number of operations required per label. For that purpose, several commands automatically mark the temporary labels into final ones, avoiding the need for the user to do that explicitly. Is up to the user to make sure that, in the end, all the necessary adjustments where in fact applied to the final label.

5.2.3   Label creation
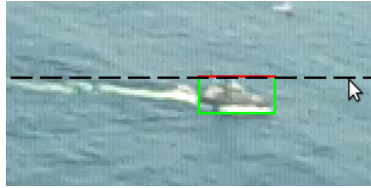
The labels can be created in two ways:
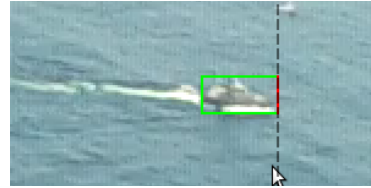
- manually - "+" key

(a) No selection (moves the whole label).

(b) Top border selected.

(c) Top border selected (mouse over horizontal straight line).

(d) Right border selected (mouse over vertical straight line).

(e) Two borders selected (mouse over the corner).

Figure 7: Ways to use the mouse.

- automatically - see 5.3

When created manually, the label is by default placed near the top left corner of the image. The user should then move the label to the right place and adjust its borders using the keyboard or the mouse.

Tip: in this case (manual creation) it is probably better to first use the mouse to do a coarse adjustment and after use the keyboard (and zoom) for fine adjustments.

5.2.4   Label deletion

Press SHIFT + "+" (i.e., "*" key) to delete the current label.

5.2.5   Label modification

Using the mouse   The mouse can be used to move the label or its borders by means of click, drag and release. The point in the image where the user clicks before dragging defines the operation realized by the mouse.

Moving the label border using the mouse   By pressing the mouse over a border or over the straight line the extends the border will that border will be moved. The selected border or borders will be shown in red color.

11

Pointing the mouse over the straight line that extends the border (but outside the border) is useful when the label is too small and is difficult to point to the border directly.

It is also possible to move two borders simultaneously by pointing the mouse at the label corner where the borders intercept.

Moving the whole label using the mouse By pressing at any point that does not select any border, the label is moved as an whole without changing its dimensions.

Moving the label border using the keyboard There are independent keys to move each of the borders and in each direction (see Figure 2). The keys disposition is organized in a logic way in what concerns the border position and the movement direction.

For example, both keys that move the left border ("a" and "s") are on the left side in relation to the other keys. In addition, the "a" key which is on the left side moves the border to the left and the "s" key which moves the border to the right is on the right side. Hopefully, this layout is intuitive enough so that the user can adjust the label without needing to look at the keyboard.

The user can even move the border faster, 10 pixels at a time, by simultaneously pressing the SHIFT key with the movement keys.

Moving the whole label using the keyboard There are no specific keys to move the label as a whole. However, that effect is easily achieved using the same keys that move the borders.

Suppose that the user intents to move the label sideways to the left. This is the same as moving both left and right borders to the left by the same ammount. We can then use the "a" and "f" keys and press them rapidly one after the other and get the movement desired. Better yet, in most cases (depending on the keyboard hardware) we can even press them simultaneously since the keyboard will still send the two keystrokes one after the other (the ordering does not matter).

This technique even works if we keep pressing the SHIFT key, allowing for large movements of the whole label, 10 pixels each time.

5.2.6 Finalizing the labels

As mentioned before, the labels are created with type equal to temporary. This always happens either for manually or for automatically created labels. Is up for the user to verify all labels, adjust as necessary and then tag the labels as final.

For work optimization purposes, it was tried to minimize the number of keys that is necessary to press for each label. Towards that purpose, some operations, namely the ones that move the label or its borders, automatically set the label type to final. It is assumed that if the operator adjusted the label then it is already final or is in the process of becoming final. It is up to the user to make sure that all necessary adjustments are in fact made.

If no adjustment whatsoever is necessary, the the label can be set to final by pressing SIFT + "t".

Pressing in "t" only (without SHIFT) sets the label as temporary again.

There is yet the ENTER key which sets the current label as final and seeks to the next frame. This combination is useful when reviewing many labels (for example, after applying the automatic methods of subsection 5.3). Probably in that case, many of the labels do not need adjustment and suffices to mark them as final and go to the next frame, and for all that the user only needs to press ENTER.

5.3 Automations

One way to alleviate the task of labeling is to use the available automations that, even though do not intend to replace the manual labeling, they can assist into making it easier. Four automatic methods are available, two of which operate image by image while the other two operate for a set of images in one step.

These methods are described next.

5.3.1 Copy

Press the "p" key.

This command copies the label of the previous frame to the current frame. The previous frame label needs to exist and be final, otherwise the command does nothing.

Assuming that the objects move very little from frame to frame, it is faster to copy the previous label and do small adjustments then to create a label from the start for every frame.

5.3.2 Copy with search

Press the "o" key.

This command is identical to the previous one but, after copying the previous frame label, also searches, in the current frame and in the neighborhood of the copied label, for the image contained in the original label. The search is based on the minimization of the quadratic error between the two images. Finally, the position of the label is adjusted to the position where the error was smaller.

In most cases this approach produces very good labels only requiring occasional small adjustments.

5.3.3 Linear interpolation

In this method two non contiguous frames are manually labeled and then linear interpolation is used to estimate the labels of the frames in the middle. See the example in Figure 8.

To use this method the following steps are required:

1. create and finalize the label in the first frame of the interval

2. create and finalize the label in the last frame of the interval

3. seek to the last frame of the interval
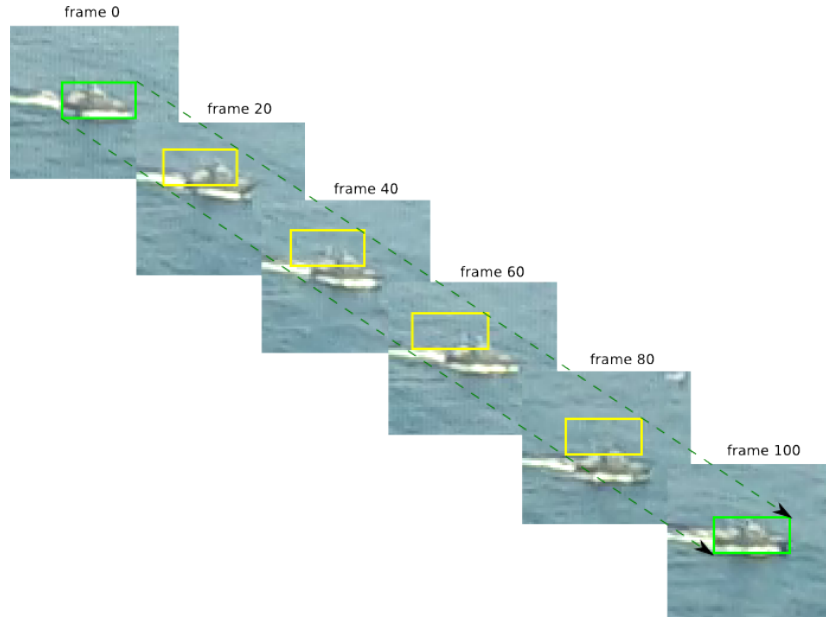
4. press the "i" key

Figure 8: Example of labels obtained automatically using linear interpolation between frames 0 and 100. The labels at the extremes (frames 0 to 100, green color) where defined manually. The others (yellow) where interpolated.

By pressing "i", the method goes backward in the video until it finds the previous final label. Then it does an linear interpolation between that label and the current frame label, either creating temporary labels for the frames in between or changing them if they already existed.

This method applies well to cases where the object moves along a line in the image. Many of the movements found are linear or can be approximated by smaller linear sections. Using this method, together with the following method (subsection 5.3.4), allows the user to automatically obtain very good approximations to the movement of the object, minimizing the number of manual adjustments necessary.

5.3.4 Search in a sequence

This method adjusts the temporary labels previously defined (for example, by the previous method - subsection 5.3.3) in a set of frames between two final labels also previously defined. See example in Figure 9.

To use this method the following steps are necessary:

1. create the final labels at the two extremes of the frame set

2. create the temporary labels inside the set

3. seek to the frame at the extreme end of the set
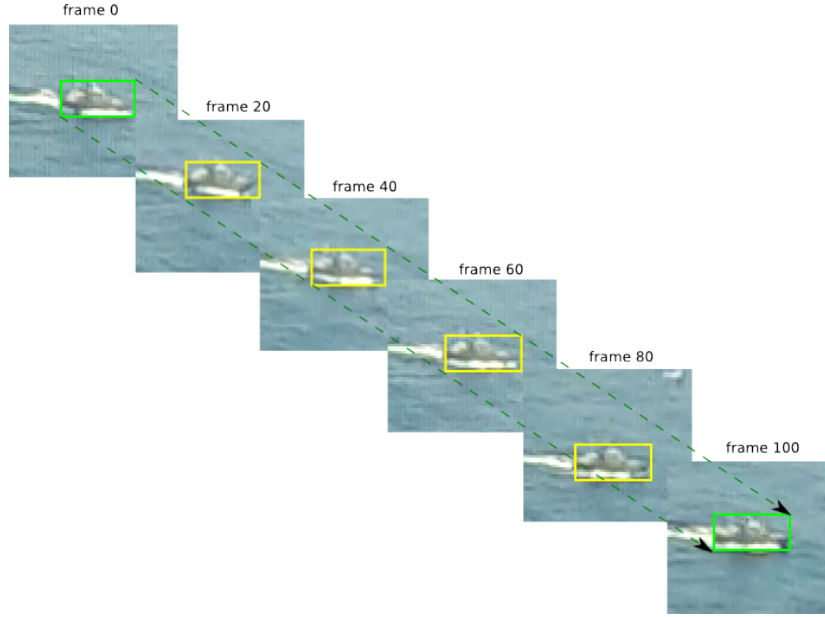
4. press the "u" key

Figure 9: Labels obtained after applying the search in a sequence method (compare with the previous figure).

By pressing "u", the method goes back in the video until it finds a final label, thus defining the frame where the set starts. Then it saves the images inside the labels at the extremes of the set.

For each frame within the set, the method searches in the neighborhood of the temporary label of this frame for the two images saved. The search is based on the minimization of the quadratic image error, similar to subsection 5.3.2, but now two errors are obtained, one for each image. The errors are weighted according to the temporal distance between the current frame and the frames of the two saved images and then the lowest one is chosen. The current temporary label is then adjusted accordingly to the position found for the image chosen.

The weighting of the error based on time distance is justified by the assumption that the closer the images, the similar they should be. Using the two images at the extremes together with this weight allows for the method to work in larger sets with larger time intervals, even if the image of the object varies in those intervals (for example, the object can rotate, change size due to perspective, illumination can change, etc.).

This method is well suited to be used with the previous method (subsection 5.3.3) since it compensates well for deviations of the object from the linear trajectory. Frequently the object's trajectory is not exactly linear. This can happen because the trajectory is in fact not linear (for example, if it is circular) but can also happen due to factors external to the object, such as camera vibrations or other camera movements.

This method was applied with great success to video sequences with up to 500 frames in a single step.

5.4 Label confirmation and verification

Note that none of this automatic methods eliminate need for the manual verification by the user (frame by frame) of the labels estimated automatically. In fact, the labeling tool marks those labels as temporary forcing the user to go by all of them to manually verify, adjust and set them as final. However, these methods obtain very good, and in many cases correct, labels for a large number of frames. By using them, the user only has to do occasional small adjustments and confirm the labels that are already correct. See how in subsections 5.2.5 and 5.2.6.

5.5 Visual Help

5.5.1 Dynamic range maximization

By pressing the "1" key the dynamic range on the luminance of the current image show on screen is maximized. This option determines the maximum and minimum values of the image's luminance and scales it so that all available dynamic range of the screen is used.

This functionality is useful when the images have very low contrast and the object is not easy to be seen.

This luminance scaling is only visual and does not affect the automatic methods of section 5.3, which continue to use the original (non scaled) luminance.

Pressing again the "1" key disables this feature.