

# Video processing

## Exercise 3: **Manipulating Frames**

---

### CONTEXT

In the context of digital multimedia a digital video makes reference to a set of images (called frames) that must be display at a given rate (frames per seconds) and that may have audio associated to them. Thus, for processing a video it is necessary to reason about the frames composing the video.

### OBJECTIVES

- Put in practice your knowledge of image manipulation for manipulating video frames.
- Reason about the structure of a video, starting with frames.
- Use a video conversion tool for transforming a video in one format file to multiple formats.

### To Do & To HAND IN

For this exercise you will need a short video composed of multiple scenes (e.g., an advertisement of a perfume brand). For simplicity you can download this video from YouTube.<sup>1</sup>

- As you already know only 3 types of videos can be used in HTML5: **.mp4**, **.ogv** and **.webm**. What are the differences of each format? In which type of situation you will privilege the use of a format over the others?
- Since none of the existing browsers support all video formats you have to convert your video into **.mp4**, **.ogv** and **.webm** for ensuring large audience. Use a video converter<sup>2</sup> and then adapt the following code in order to display your video in all browsers.

```
<!DOCTYPE html>
<html>
<head></head>
<body>
  <div>
    <video id="theVideo" height="720" width="400" autoplay loop controls >
      <source src="Dior.mp4">
      <source src="Dior.ogv">
      <source src="Dior.webm">
    </video>
  </div>
</body>
</html>
```

---

<sup>1</sup> You can download videos from YouTube using the following web tool: <http://savemedia.com>.

<sup>2</sup> <http://www.mirovideoconverter.com>

- Similar to images, video frames can be manipulated at pixel level using a canvas. Reuse the functions you defined in TP2 and apply them to your video. Use the following code as a starting point.

```

<!DOCTYPE html>
<html>
<head>
  <script type="text/javascript">

    window.addEventListener('load', onWindowLoaded, false);

    var video, canvas, context;
    var videoFilter = "none";

    function onWindowLoaded() {
      video = document.getElementById("theVideo");
      canvas = document.getElementById("theCanvas");
      context = canvas.getContext("2d");
      setInterval(drawFrame, 33);
    }

    function drawFrame () {

      if(videoFilter == "inverse") {
        context.drawImage(video, 0, 0, 720, 400);
        inverseFrame();
      }

      if(videoFilter == "none") {
        context.drawImage(video, 0, 0, 720, 400);
      }

    }

    function setFilter(filter) {
      videoFilter = filter;
    }

  </script>
</head>
<body>

  <div id="Buttons">
    Filter
    <button onclick="setFilter('none');">None</button>
    <button onclick="setFilter('inverse');">Inverse</button>
  </div>

  <div>
    <canvas id="theCanvas" width="720" height="400">
      Your browser does not support HTML5 Canvas.
    </canvas>
  </div>

  <div>
    <video id="theVideo" height="720" width="400" autoplay loop controls style="display:none">
      <source src="Dior.mp4">
      <source src="Dior.ogv">
      <source src="Dior.webm">
    </video>
  </div>

```

```
</body>  
</html>
```

- As you can see from the previous step not all image operations can be applied to the frames of a video without side effects (e.g., the blur function delays the video). Describe a strategy for dealing with this kind of problems and give an example.
- *Subliminal stimuli are any sensory stimuli below an individual's threshold for conscious perception. Visual stimuli may be quickly flashed before an individual can process them, or flashed and then masked, thereby interrupting the processing*<sup>3</sup>. Assuming a frame rate of 24 fps, propose and implement a video manipulation strategy for adding subliminal images within your video. Note that they should be inserted within the frames and their duration should be short enough so that the eye cannot “consciously” detect them.

**Bonus question: To be answered completely and to be partially implemented.**

The information contained in a video is not restricted to the sequence of images (frames) composing the video; meta-data can be associated to it. A particular type of meta-data concerns the partitioning of a video in sub-sequences of contiguous images according to a semantic criterion that leads to tree like structures. For instance, in classic cinematography, films are divided into *sequences*, each composed of *scenes* that are themselves divided into *plans*, which are also composed of images.

There are particular types of meta-data called annotations that can be associated to an image or video segment (e.g., key words). Using the notion of meta-data and annotation it is possible to define a model that integrates them in an orthogonal way meaning that each structure element can be annotated independently. This can help to add information that can be true for an image but not for a plan. For example, the fact that two characters interact in a scene is not necessarily true in every image of the scene.

Since video is associated to time, it is possible to use the notion of *timeline* and *granularity* to propose a model for representing the organization of a video content. This can help to model its sequential structure (e.g., the number of reference points in a time line is equal to the number of images of the video) and the organization levels using different temporal granularities (e.g., point and interval).

According to this description, you should propose:

1. A model for representing the structure and the content of a video (use the notion of meta-data and annotation). You should include the temporal aspect but then propose any other concepts for representing the structure and content of the video. How can you implement it in order to provide “enriched” videos that can be then manipulated according to their content (e.g., retrieve the images that appear within second 15 and second 20)?
2. Using your model propose a strategy to generate video summaries. Implement it.
3. Propose a strategy for retrieving sub-sequences of a video for example:
  - a. Give the names of the characters that appear at least once during the first seconds of a video assuming that the presentation rate is 30 images/second.
  - b. Give the scenes where a character appears in at least half of the images of a scene.

---

<sup>3</sup> [http://en.wikipedia.org/wiki/Subliminal\\_stimuli](http://en.wikipedia.org/wiki/Subliminal_stimuli)