Review of Bachelor's Thesis

Student: Štafl Roman
Title: Video Summary (id 24930)
Reviewer: Herout Adam, prof. Ing., Ph.D., DCGM FIT BUT

1. Assignment complexity — less demanding assignment

The assignment per se belongs among the more challenging ones. However, the approach that the student has taken is trivial.

2. Completeness of assignment requirements — assignment fulfilled with minor reservations

The student proposed "some" system for video summarization.
The way it was designed is very basic and the tech report does not discuss any improvements or iterations - it appears that some solution occurred to the student so he implemented it and that was the end of the project. Section 4.4.1 essentially declares that the solution is practically unusable, the experimental evaluation is also not making much sense.

3. Length of technical report — within minimum requirements

The extent of the tech report is on the bottom limit.

4. Presentation level of technical report — 55 p. (E)

Section 2 with the "theoretical background" contains basic statements (video is composed of scenes and shots - this is not even true for surveillance videos on which the work was focused; illustrative figures such as 2.3, 2.4, 2.15 etc.; historical and trivial approaches to extraction of image features; trivial metrics of feature vector distances, sec 2.5).
Section 3 with the system design again contains trivial illustrations (fig 3.1), and rather vague system description (there is zero equations in the section). It does not contain any reasoning why the selection of the feature extractor (HOG), or similarity matrix (fig 3.3), or summed vector across rows of this matrix (fig 3.4) should be in any way a good approach to video summarization.
Section 4 about the implementation shows that individual parts of the solution are interconnected through pickle files. The text does not mention any iterative improvement of the originally proposed algorithm. Reflection how good the achieved summarization is, is missing. It is not clear how the evaluation tools extract any information about the "important" parts of the videos that should be kept in the summarization.
All graphs with the PRC curves indicate that the system DOES NOT work at all (without the author being able to reflect on that): The higher number of included video fractions, the better recall, but the precision is not changing measurably. The graph in fig 4.5 does not make ANY sense - including 50 best snippets destroys the precision to a very low number, but including best 100, 200, 350 of the same snippets returns the precision to a great value of 0.9, which does not change further - the graphs appear to be plotted nonsense data that the student was not able to comprehend and interpret.

5. Formal aspects of technical report — 55 p. (E)

The text is written in good and understandable English and most of the typesetting is average.
Then the text includes images with very strong lossy compression (fig 2.8 and others) and mostly "equations" included as images with lossy compression and captioned Figure X (Fig 2.12-2.14). For someone who applies to receiving a university degree in the field of image processing, this is a sign of astonishing ignorance.

6. Literature usage — 65 p. (D)

The text refers to a high number of fairly sound sources. The shape of the program and tech report do not indicate real understanding of the ideas contained in the sources. The approaches used in the work are fairly retro - today's feature extractors and tools for video processing are based on considerably different approaches, which the tech report and the solution itself completely ignore.

7. Implementation results — 50 p. (E)

The student proposed and programmed some solution of the assignment and carried out some experiments. The solution is composed of a handful of python scripts, which have typically literally a few lines. Individual scripts are exchanging data only by using files whose names are given by literals in the source code. Commands for plotting the graphs (probably used in the tech report) are directly placed in the code, so the plotting seems to be an inevitable part of the "production" processing. Overall, the quality of the software engineering is very low. The extent of the created work and its quality in terms of results and selected techniques is very low.

8. Utilizability of results

No.

9. Questions for defence

- Show an example of a summarized video, which your system created and that will show that the summarization actually works.
- Explain curves from figs. 4.1, 4.3, 4.5, 4.6.
- How to interpret values from the graph in fig. 3.4 and 4.8? Why many of the values are 0, while others are overall equal? Why should one think that the numbers plotted in this graph have any relation to parts of video important in the summary?

10. Total assessment   **55 p. sufficient (E)**

The assignment is comparatively more difficult.

The created solution is extremely questionable and unconvincing.

The tech report suffers from very vivid errors (equations inserted as images with lossy compression captioned Figure X; reprinted illustrative trivial images; strange and uninterpreted PRC curves; etc.).

In Brno 2 June 2022

Herout Adam, prof. Ing., Ph.D.
reviewer