ISSN (Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 8, August 2014

Accelerating video frames classification with metric based scene segmentation

Adam Blokus, Jan Cychnerski, Adam Brzeski

Department of Computer Architecture, Gdańsk University of Technology, Poland

ABSTRACT: This paper addresses the problem of the efficient classification of images in a video stream in cases, where all of the video has to be labeled. Realizing the similarity of consecutive frames, we introduce a set of simple metrics to measure that similarity.

To use these observations for decreasing the number of necessary classifications, we propose a scene segmentation algorithm. Performed experiments have evaluated the acquired scene sizes and classification accuracy resulting from the usage of different similarity metrics with our algorithm. As a result, we have identified those metrics from the considered set, which show the best characteristics for usage in scene segmentation.

KEYWORDS: scene segmentation, image distance, image classification

I. Introduction

Medical imaging techniques see ongoing improvements, among others the development of new video data acquisition methods (e.g. Wireless Capsule Endoscopy - WCE [1]) and a constant acceleration of processing acquired data. In these fields there are many issues related to gathering and processing large amounts of video data. For example, a single WCE video can be up to 8h long and require 2-3h of time of a medical specialist to examine [2].

Depending on the particular application, it may be required to provide a general conclusions about the whole video or a detailed analysis of each separate frame. The former case is related to diagnostics, where the most important information is whether the patient has any internal bleedings or lesions, and any additional information about them. The latter case may be considered when creating a database for image recognition algorithms or educational purposes - when the final user (here: a person or algorithm) operates on various pictures, comparing cases of diseases with healthy tissues.

In real life videos, especially when a sufficiently high frame rate is assured, consecutive frames are very similar, and present just minor modifications to the adjacent frames. In medical applications that means that the recorded stream can be perceived as mostly "continuous" [3], with objects entering and leaving the scene gradually [4].

The continuous character of videos allows to consider a division into scenes [5] - shorter sequences of similar frames, sharing a given label. Methods for finding such divisions have been developed especially for live action films [6], [7], where scene and shot changes are instant and significantly easier to spot by the rapid change of frame characteristics.

In medical imaging a popular approach involves the detection of characteristic or specific frames to acquire a so called *video summary* [8], [9], [10], [11]. These frames are representatives of their direct neighborhoods, with implicitly assumed changes of scenes between the.

The pace of the change of the video is measured by metrics specifying the relative difference of consecutive frames. Such functions and their properties have been defined in section II.

In this paper we will present the results of evaluating a new scene segmentation algorithm incorporating three selected metrics with different parameter sets. The evaluation is performed in terms of the accuracy of the final classification of all scenes as a whole. The whole procedure has been further described in section III, whereas its results can be found in section IV.

ISSN (Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 8, August 2014

II. Preliminaries

A. Image metrics

To determine how much two frames differ and how much the view in the video changes, a function for comparing frames has to be defined. Such functions will further be called **metrics**. Our definition [1] of a metric d requires it to fulfill the following properties:

$$d(a,b) \geq 0 \tag{1}$$

$$d(a,a) = 0 \tag{2}$$

$$d(a,b) = d(b,a) \tag{3}$$

$$d(a,b) < d(a,c) + d(c,b) \tag{4}$$

In the considered usage the value of the metric is expected to quantify the visual similarity of images in some way.

The definition above allows for any non-negative metric values. It is worth noting though, that the set of all possible images remains finite[2], which implies the existence of an upper boundary M of the image distance. For an easier comparison and evaluation, the metric values are linearly normalized to the range $[0;1]$ (by dividing the value of a metric by upper bound M for the considered image size).

For our purposes we have chosen to consider the following metrics:

- Simple distance (SD) - the l_1 distance between the images (vectors of pixel values)
- Simple distance on processed image (SP) - the Simple distance of two images after bluring and downscaling.
- Histogram distance with k bins (HD, HD$_k$) - the l_1 distance between k-bin HSV color histograms of the images.

Algorithm 1 Scene segmentation

Input:
  $F[]$ - sequence of N frames
  T - threshold value for metric d
Output: $S[]$ - scene assignments
  scene $\leftarrow 1$, *current* $\leftarrow 1$, *next* $\leftarrow 1$
  while *current* $< N$ **do**
    #Expansion
    while $d(F[current], F[next]) < T$ **and** *next* $< N$ **do**
      $S[next] = scene$
      $next \leftarrow next + 1$
    end while
    $scene \leftarrow scene + 1$
    $mid \leftarrow next$
    #Reduction
    while $d(F[mid], F[next]) < d(F[mid], F[current])$ **do**
      $S[mid] = scene$
      $mid \leftarrow mid - 1$
    end while
    $current \leftarrow next$
  end while

[1]Strictly speaking, the presented conditions define a *pseudometric* on the image space, since it is possible that $d(a,b) = 0$ for some $a \neq b$. The considered functions are proper distance functions only for the processed vectors acquired in a way defined for each metric.

[2]The number of all possible pixel values for a given resolution and color depth is finite.

ISSN (Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 8, August 2014

B. The algorithm

In this section we propose a simple algorithm for scene segmentation of (mostly) continuous videos. For each scene a specific frame which defines it is chosen. Next, the scene is created in two main steps:

- **Expansion** - consecutive frames are assigned to the scene until a frame is reached whose difference from the specific frame exceeds a given threshold. That frame will be the next specific frame.
- **Reduction** - frames from the end of the current scene, which are more similar to the next specific frame than to the current one are reassigned to the next scene.

The pseudocode of the algorithm has been presented in Algorithm 1 (exact with respect to treating some boundary cases).

III. Experiments

Performed experiments involved the evaluation of the algorithm on a set of exemplary recordings from endoscopic examinations. Six representative films have been chosen, fulfilling the following criteria:

- at least 1000 frames long
- the recognized property is present in between 20% to 80% of the frames
- there are at least 5 changes of the recognition status in the video

Those requirements have been set to prevent overrating algorithms just propagating a single result on all frames and evaluating algorithms on rare "chaotic" videos.

The scene segmentation algorithm has been applied to every film with five different metrics, using the thresholds listed in Table I. Different threshold sets are a result of different characteristics of the metrics - with SD and SP showing larger changes in value.

Metric	Thresholds
SD	0.4, 0.42, 0.44, 0.46, 0.48
SP	0.4, 0.42, 0.44, 0.46, 0.48
HD_4	0.05, 0.1, 0.15, 0.2, 0.25, 0.3
HD_6	0.05, 0.1, 0.15, 0.2, 0.25, 0.3
HD_8	0.05, 0.1, 0.15, 0.2, 0.25, 0.3

Table I: Parameter sets

IV. Results

The first observation of the evaluation is the relatively high distance in SD and SP metrics for seemingly similar images. This observation is related to both these metrics being sensitive to even minor shifts and allowing to detect images with a very high amount of common static areas.

For various tested threshold values the average scene lengths have been computed and the results presented in Table II and Table III. As expected, a clear positive correlation between the threshold value and scene length can be seen for all metrics.

k \ T	0.05	0.1	0.15	0.2	0.25	0.3
4	3.2	6.2	10.0	15.3	21.9	29.8
6	2.8	5.0	8.1	11.8	16.7	23.1
8	2.5	4.5	7.2	10.1	14.2	19.9

Table II: Average scene lengths for HD_k metrics

Figure 1 presents the relation of recognition accuracy and metric thresholds. A negative correlation can be seen between those two values. Results for the H_8 metric, which got the best results of the H_k metrics, show that an

ISSN (Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 8, August 2014

T	0.4	0.42	0.44	0.46	0.48
SD	32.6	73.3	91.3	131.8	177.8
SP	16.3	23.1	32.1	48.2	57.7

Table III: Average scene lengths for SD and SP metrics

accepted change of 25% of the possible distance (or: increasing the amount of differences six times) results in a change of less than 10% in accuracy.

Since different threshold values have been tested for the two groups of metrics, it is important to note, that the values between the groups are not comparable in this graph.

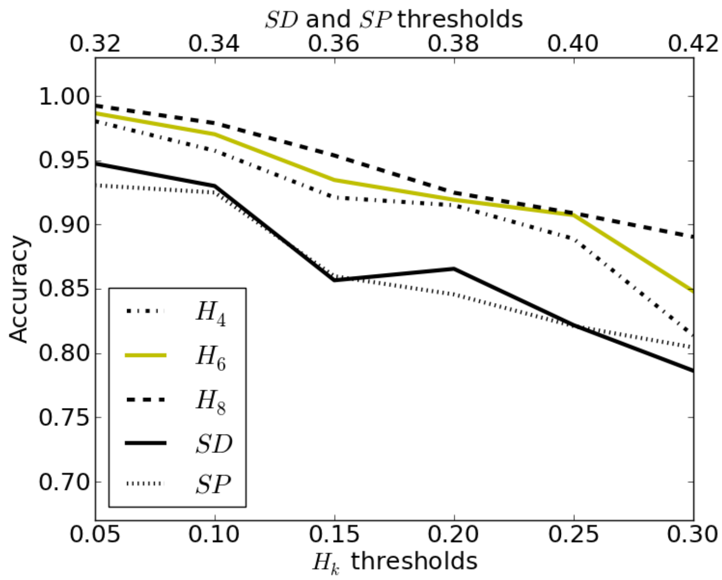

Figure 1: Accuracy change with threshold

The graph in Figure 2 presents the relation between the acquired average scene sizes and the resulting classification accuracy. It can be seen that all of the H_k metrics acquired similar results and outperformed the SD and SP metrics. This shows that the scene segmentation algorithm with the H_k metrics acquires a better division into scenes and assignment of their specific frames.

High accuracy values of over 95% are preserved for scene sizes of up to six frames. With such results. costly recognition algorithms might be improved to operate on whole scenes, with a scene segmentation algorithm tuned in respect to given time limitations depending on an accepted performance/accuracy tradeoff.

V. SUMMARY

In this paper a new method for accelerating the classification of frames in video sequences has been shown. Preserving high reliability, the number of frames to be processed by a classifying algorithm can be decreased by a factor of over 80%. Three general types of metrics have been evaluated, with shift-insensitive metrics based on HSV histograms performing simple image distances.

A broad range of possible metric definitions and parametrizations leaves an open space for further experiments in this subject.

ISSN (Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 8, August 2014

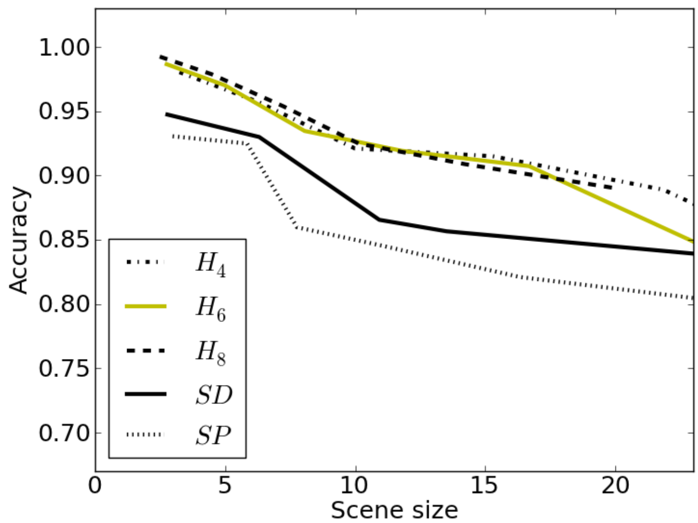

Figure 2: Accuracy change with scene size

REFERENCES

1. Alexandros Karargyris and Nikolaos Bourbakis. Wireless Capsule Endoscopy and Endoscopic Imaging: A Survey on Various Methodologies Presented. *IEEE Engineering in Medicine and Biology Magazine*, 29(1):72–83, 2010.

2. Michael Liedlgruber and Andreas Uhl. Computer-aided decision support systems for endoscopy in the gastrointestinal tract: a review. *IEEE reviews in biomedical engineering*, 4:73–88, 2011.

3. Zdzislaw Pawlak. On Some Issues Connected With Roughly Continuous Functions. 1995.

4. Hai Vu, Tomio Echigo, Ryusuke Sagawa, Keiko Yagi, Masatsugu Shiba, Kazuhide Higuchi, Tetsuo Arakawa, and Yasushi Yagi. Contraction detection in small bowel from an image sequence of wireless capsule endoscopy. *Medical Image Computing and Computer-Assisted Intervention*, 10(Pt 1):775–783, 2007.

5. John M Gauch, Susan Gauch, Sylvain Bouix, and Xiaolan Zhu. Real time video scene detection and classi T« cation. *Wall Street Journal*, 35:381–400, 1999.

6. Tong Lin and Hong-jiang Zhang. Automatic video scene extraction by shot grouping. *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*, 4:39–42, 2000.

7. Ufuk Sakarya and Ziya Telatar. Video scene detection using graph-based representations. *Signal Processing: Image Communication*, 25(10):774–783, November 2010.

8. That Mon Htwe, Chee Khun Poh, Liyuan Li, Jiang Liu, Eng Hui Ong, and Khek Yu Ho. Vision-based techniques for efficient Wireless Capsule Endoscopy examination. In *2011 Defense Science Research Conference and Expo DSR*, pages 1–4. Department of Computer Vision and Image Understanding, Institute for Infocomm Research, Singapore 138632, IEEE, 2011.

9. Selen Atasoy, Diana Mateus, Joe Lallemand, Alexander Meining, Guang-Zhong Yang, and Nassir Navab. Endoscopic video manifolds. *Medical Image Computing and Computer-Assisted Intervention*, 13(Pt 2):437–445, 2010.

10. Giovanni Gallo, Eliana Granata, and Alessandro Torrisi. Information Theory Based WCE Video Summarization. *Pattern Recognition ICPR 2010 20th International Conference on*, 0:4198–4201, 2010.

11. Giovanni Gallo and Alessandro Torrisi. Boosted Wireless Capsule Endoscopy Frames Classification. In *PATTERNS 2011, The Third International Conferences on Pervasive Patterns and Applications*, pages 25–30, 2011.