

Scene Change Detection for Uncompressed Video

Patrick Seeling

Department of Computing and New Media Technologies

University of Wisconsin – Stevens Point

Stevens Point, Wisconsin 54481-3871

Email: pseeling@ieee.org

Abstract—In this paper, we present a scene change or shot boundary detection method based in the changes in entropy of differences between uncompressed video frames. As in the uncompressed domain, cues for scene or shot boundaries are not available, detecting video content features is a non-trivial and typically requires additional complexity in the evaluation. The entropy presents a metric for the complexity of information. Used on the differences between video frames, the entropy is able to measure the complexity of changes. We find that due to content dependency, however, the relative entropy changes in the sequence of video frames is a better indicator for detection.

An evaluation of the presented approach finds that detection for a combination of video test sequences can be reliably performed using the U component of uncompressed YUV 4:2:2 video only.

I. INTRODUCTION AND RELATED WORK

The domain of video annotation, indexing, and retrieval has attracted a large body of research in the past. One of the main drivers for this area of research is the readily available computing power in normal desktop computers, digital video equipment, and motivation to share content over the Internet in various forms of access models, i.e., sharing via web pages, live streaming, and so on. It is expected that the amount of digital video that is available will continue to grow. With this vast amount of data available, annotation and logical indexing of video becomes a desired feature. One standard for the annotation of multimedia data is MPEG-7, see, e.g., [1]. Amongst different approaches to annotation, segmenting the video into its scenes or shots is the most intuitive and basic. Several different approaches exist to detect shot boundaries in video, see, e.g., [2], [3] for an extensive overview of different algorithms and their classifications.

Reliable and universal detection of scene changes or shot boundaries in the uncompressed domain in a fast manner, however, is still a challenge. While in the compressed video domain, typically cues can be obtained that are readily available from the encoded video stream (e.g., motion vector intensity or transform coefficients), these cues are not available in the uncompressed domain. To alleviate this problem, the authors in [4] employ a two-stage histogram-based method to determine scene changes and filter out unwanted false detections. More sophisticated approaches have to implement additional methods, such as edge detection, see, e.g., [5]. Other venues of research employ neural networks to determine scene changes based on several pixel- and color-based features of uncompressed video, see, e.g., [6]. In [7], the authors apply

fuzzy logic approaches to detect scenes. Recently, the authors of [8] employed an autoregressive model based on the color histogram in the uncompressed domain to determine scene changes.

In the following, we present a scene detection algorithm that is solely based on the entropy of differences between frames. The motivation for using this approach is that with changes between frames, the entropy of the differences will increase. We evaluate the algorithm against a short video test sequence and a *Combined* video sequence, which is derived from multiple test sequences. We present the performance metrics precision and recall obtained with the proposed algorithm before we conclude.

II. ENTROPY OF FRAME DIFFERENCES

The scene boundary detection method introduced in the following is applied on uncompressed video frames using the YUV 4:2:2 format. This format is typically used for video coding and transcoding for a variety of video codecs. The YUV format describes each individual pixel by its luminance (Y) and two color differences (U) and (V), also known as chrominance components.

We denote the i -th byte in the n -th frame out of $n = 0, 1, \dots, N$ frames as $F_n(i)$. Furthermore, let $Y_n(i_Y)$ denote the luminance byte values and $U_n(i_{UV}), V_n(i_{UV})$ denote the i -th byte values for the two chrominance components. Note that due to the chroma sub-sampling, $0 \leq i_{UV} \leq \frac{i_Y}{4}$, i.e., the two chrominance components are restricted to half the resolution of the luminance component. Sub-sampling is used since the human eye is most sensitive to the luminance component. In the 4:2:2 format, the U and V values are each sub-sampled for a group of 4 luminance pixels and stored by component in a grouped manner on disk. The manner in which the individual values are sub-sampled and typically stored on disk is illustrated in Figure 1 for an individual frame.

We denote the probability for a specific byte value $F_n(i)$ in frame n as p_{F_n} with $F_n(i) \in \{0, \dots, 255\}, \forall i$. The entropy gives a measure for the complexity of an individual frame's information content. The entropy for the byte values of frame n is calculated as in Equation 1.

$$H_n = - \sum_{b=0}^{255} p_{F_n} \cdot \log_{256}(p_{F_n}) \quad (1)$$

The entropies for a frame's three individual components can be calculated in a similar manner.

In order to derive a better estimator for the changes between frames, we calculate the entropy for the differences between frames n and $n - 1$ for all frames $n \geq 1$. The differences between two complete frames are calculated as in Equation 2.

$$F_{n,n-1}(i) = |F_n(i) - F_{n-1}(i)|_{\forall i} \quad (2)$$

The entropy for the frame differences is then calculated as in Equation 1 for $p_{F_{n,n-1}}$. The entropy of the difference frames (or the differences frames' components) can then be used to determine the scene boundaries by comparing subsequent difference frame entropies. We note that this approach requires at least three frames to be processed, as illustrated in Figure 2.

III. SCENE CHANGE DETECTION METHOD

The initial evaluation of the algorithm presented in this paper is performed on the *News* sequence in the QCIF format ($i = 38016, i_Y = 25344, i_{UV} = 6336$). This sequence features two news anchors and a changing background with varying displays of dancers, as illustrated in Figure 3. Details for the *News* video sequence's content are provided in Table I. We illustrate the entropy for the frame differences in Figure 4. We observe that the entropy of frame differences "spikes" where the content of the underlying video sequence changes at frames 91, 262, and 241. This "spike" of the entropy can be used to detect the changes in scene content.

For the evaluation of the scene change algorithm introduced in Section III with respect to shot boundary detection, we now employ a *Combined* video sequence in the QCIF format, which is derived by concatenation of multiple video sequences. The sequences' details are given in Table I. The resulting entropy for the differences between complete frames $F_{n,n-1}$ is illustrated in Figure 5. We initially observe that the changes between individual scenes are represented by spikes in the entropy of the differences between full frames. We additionally observe that for each original sequence, a separate level and behavior of the entropy of frame differences can be observed in Figure 5. The scenes containing more motion and camera movement exhibit more varying entropies. For the *Husky* sequence, a generally high level of entropy is observed, while the *Bow* sequence's entropy exhibits more pronounced changes or "spikes" of the entropy.

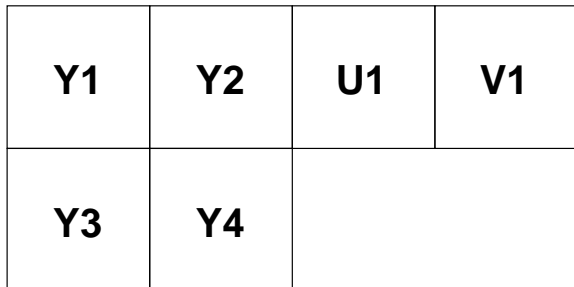


Fig. 1. YUV 4:2:2 pixel format and single video frame storage.

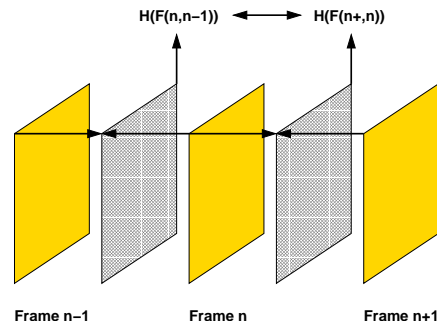


Fig. 2. Calculation of difference frames and entropy values.

Sequence	Frame	Content
<i>News</i>	0	A news sequence with two anchors and varying background.
	91	Background is one dancer close-up.
	151	Background is two dancers.
	241	Background is one dancer close-up.
<i>Salesman</i>	301	A salesman presents his product.
<i>Akiyo</i>	750	A news anchor talks in front of a static background.
<i>Husky</i>	1050	Several runners and their dogs, with some camera panning and zooming.
<i>Bow</i>	1300	A person enters, bows, and leaves.
<i>Hall Monitor</i>	1600	A hall monitor with people passing by.

TABLE I
DETAILS FOR THE CONCATENATED *Combined* VIDEO SEQUENCE.

The entropy of the differences for the individual frame components is illustrated in Figure 6. We observe that overall, all three individual components exhibit similar characteristics of the entropy compared to the complete frame. We note, however, that the two chrominance components exhibit a significantly lower level in their entropy values than the luminance component for the *Husky* and *Bow* sequences. We conclude that comparing the entropies of the difference frames for identification of scene boundaries itself is not advisable with respect to the different levels of entropy due to the content differences.

When the change in the entropies of the frame differences is evaluated, however, the scene changes can be detected



Fig. 3. Example screenshot of the *News* video sequence.

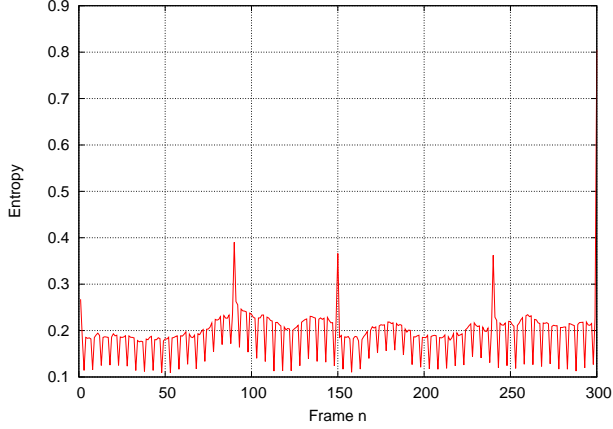


Fig. 4. Entropy of the frame differences for the *News* QCIF video sequence.

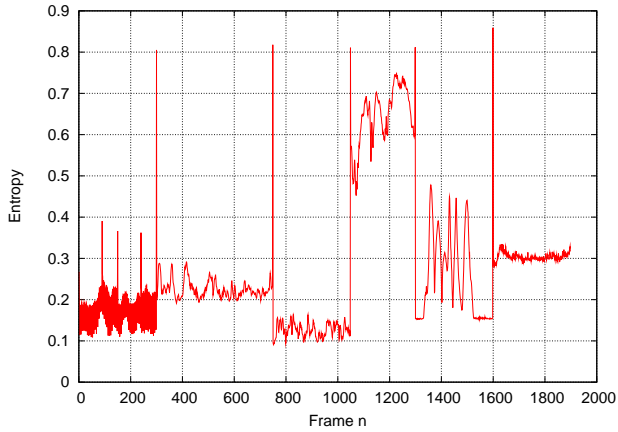


Fig. 5. Entropy of the frame differences for the *Combined* QCIF video sequence.

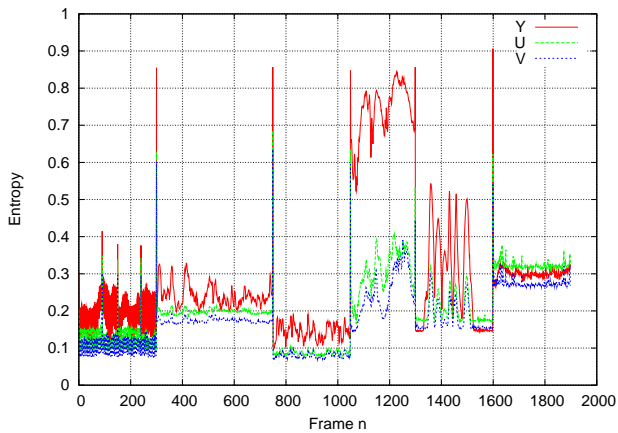


Fig. 6. Entropy of the frame differences for the individual Y,U, and V components of the *Combined* QCIF video sequence.

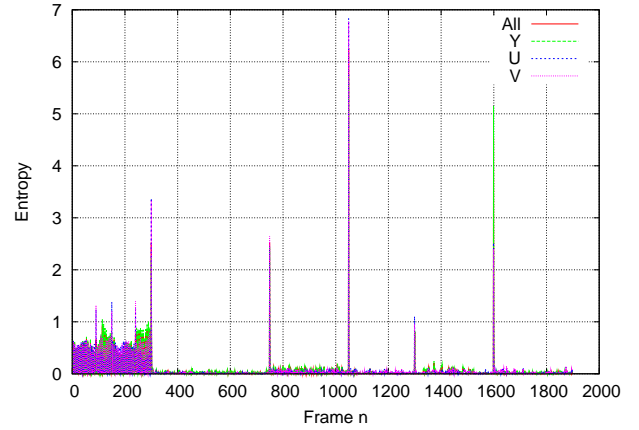


Fig. 7. Relative entropy changes for the frame differences of the *Combined* QCIF video sequence.

in a manner that takes the underlying content into account. As illustrated in Figures 5 and 6, the level of the entropy changes with the scene’s content. The relative changes in the entropy values are hence a better estimator to determine scene boundaries. We calculate the relative entropy changes as

$$H_{n,n-1} = \frac{|H_n - H_{n-1}|}{H_{n-1}}. \quad (3)$$

The resulting relative entropy changes are illustrated for *Combined* video sequence in Figure 7. We initially observe that for the first part of the *Combined* video sequence, a fairly high level of noisy changes in the relative entropy can be examined, whereas this behavior is not visible for the other parts of the sequence. This behaviour can be explained by the background video in the *News* video sequence, see Figure 3. Thus, this behavior is content-dependent. Secondly, a closer examination of the “spikes” in the relative entropy yields that they occur at the scene change boundaries. We also note that these “spikes” are at least in the region of 100 percent changes in the relative entropy. This leads to the conclusion that changes of more than 100 percent in the relative entropy are indicators of scene changes.

To detect scene changes following this approach, we define the threshold for the detection as

$$H_{n,n-1} \geq t. \quad (4)$$

IV. PERFORMANCE EVALUATION

Typical performance metrics for the detection of shot and scene boundaries are *Recall* and *Precision*. The recall value presents a measure for the correct detection of changes, whereas the precision measures the correctness of detected changes. Let D denote the correct number of detections, D_F denote the number of false detections, and D_M denote the number of missed detections. Recall and precision can then

TABLE II

PRECISION AND RECALL VALUES FOR DIFFERENT THRESHOLDS FOR THE *Combined* VIDEO SEQUENCE.

Thres. t	All		Y		U		V	
	Prec.	Rec.	Prec.	Rec.	Prec.	Rec.	Prec.	Rec.
1	1	0.5	1	0.875	1	1	1	0.875
1.1	1	0.5	1	0.875	1	1	1	0.875
1.25	1	0.5	1	0.75	1	0.75	1	0.75
1.5	1	0.5	1	0.5	1	0.5	1	0.5
2	1	0.5	1	0.5	1	0.5	1	0.5

be calculated as

$$\text{Recall} = \frac{D}{D + D_M} \quad (5)$$

$$\text{Precision} = \frac{D}{D + D_F} \quad (6)$$

For the review of the proposed method on the *Combined* video sequence, we provide the values for precision and recall in Table IV for different thresholds t of the relative entropy. We observe that precision for the entropy of the complete frame differences is high for all components combined and individually. We furthermore note that there is no impact of different threshold levels on the precision. For the recall values, on the other hand, we observe a low value for the complete frame's entropy and slightly higher values for the Y and V components. Overall, only the U component yields complete detection of all scene changes with a threshold level close to 1.

An additional observation from this characteristic is the time required to parse the complete video, especially if resolutions above QCIF are considered. As the U component's size in bytes is only $\frac{1}{6}$ of the complete video frame's size, the scene detection speed can be greatly increased.

V. CONCLUSION AND OUTLOOK

In this paper, we presented a method to detect scene changes or shots based on the relative changes in the entropy of difference frames in the uncompressed domain. As the differences between frames increase, the entropy will increase as well. However, due to content dependency of the level and behavior of the entropy, we use the relative changes in the entropy of differences between frames for detection. We found that based on the performance metrics precision and recall, the fastest and most reliable approach to detect scene changes or shots is to calculate the relative entropy differences for the U component only.

Future research venues will include evaluation of the detection method presented herein for a larger variety of video content and additional refinements to detect additional changes in the underlying video based on the entropy of the frame differences.

REFERENCES

- [1] P. Salembier and J. Smith, "Mpeg-7 multimedia description schemes," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 6, pp. 748–759, Jun. 2001.
- [2] R. Lienhardt, "Reliable transition detection in videos: A survey and practitioner's guide," *International Journal of Image and Graphics (IJIG)*, vol. 1, no. 3, pp. 469–486, 2001.
- [3] C. Cotsaces, N. Nikolaidis, and I. Pitas, "Video shot detection and condensed representation," *IEEE Signal Processing Magazine*, vol. 23, no. 2, pp. 28–37, Mar. 2006.
- [4] X. Yi and N. Ling, "Fast pixel-based video scene change detection," in *Proc. of the IEEE International Symposium on Circuits and Systems (ISCAS)*, vol. 4, Kobe, Japan, May 2005, pp. 3443–3446.
- [5] R. Zabih, J. Miller, and K. Mai, "A feature-based algorithm for detecting and classification of production effects," *ACM Multimedia Systems*, vol. 7, no. 1, pp. 119–128, Jan. 1999.
- [6] M.-H. Leea, H.-W. Yoob, and D.-S. Jang, "Video scene change detection using neural network: Improved art2," *Expert Systems with Applications*, vol. 31, no. 1, pp. 13–25, Jul. 2006.
- [7] H. Fanga, J. Jiang, and Y. Feng, "A fuzzy logic approach for detection of video shot boundaries," *Pattern Recognition*, vol. 39, no. 11, pp. 2092–2100, Nov. 2006.
- [8] W. Chen and Y.-J. Zhang, "Parametric model for video content analysis," *Pattern Recognition Letters*, vol. 29, no. 3, pp. 181–191, Feb. 2008.