AUTOMATIC QUALITY ANALYSIS FOR FILM AND VIDEO RESTORATION

Peter Schallauer, Werner Bailer, Roland Mörzinger, Hermann Fürntratt, Georg Thallinger

JOANNEUM RESEARCH, Institute of Information Systems & Information Management, Graz, Austria
{first name.last name}@joanneum.at

ABSTRACT

A considerable amount of work in larger film and video restoration projects is dedicated to manually exploring the audiovisual content in order to estimate the costs for restoration and to plan the restoration. Manual exploration is a significant cost factor. In this paper we propose automatic content analysis algorithms and summarization techniques which allow the reduction of manual inspection time in a software based restoration environment. The throughput requirement for analysis of dust and other defects is reached by sparse application of the detectors in the image sequence while retaining sufficient detection accuracy. Analysis result metadata are represented in a MPEG-7 standard compliant way. The proposed defect summary visualization tools facilitate efficient exploration of visually impaired content by the user.

Index Terms — image sequence analysis, image sequence restoration, visual quality, MPEG-7, summarization

1. INTRODUCTION

Automatic quality analysis of audiovisual content is an important tool in several steps of the media production, delivery and archiving process.

Broadcasters are checking audio and video quality within the ingest process, after editing, after encoding and before play-out for terrestrial, satellite and cable broadcast or for delivery to internet and video-on-demand services. Archives are checking for content integrity at archive ingest and delivery. Content providers are checking post production content for correct encoding and conformance to the required quality and format standard before dispatching to the broadcasters or other service providers. These use cases have in common that mainly technical properties of the material are checked, e.g. stream compliance, GOP structure, playtime, aspect ratio, resolution or MXF compliance. Additionally some content properties are checked, e.g. blocking, luma/chroma violation or noise.

For restoration the results of automatic quality analysis aim at improving the efficiency of restoration cost estimation and restoration planning and at supporting manual restoration.

Restoration as an application of quality analysis has almost not been investigated. Previous work on supporting a real-time restoration process by on-line, software based content analysis can be found in [3]. Due to the real-time requirement of that approach only limited content analysis functionality can be implemented. As opposed to on-line content analysis, our work focuses on off-line quality analysis.

Quality analysis in the software-based restoration process faces several challenges, e.g. concerning the requirements to analysis algorithms. Section 2 outlines these requirements and presents how existing defect detection algorithms can be modified to fulfill them. In Section 3 the modified algorithms are evaluated.

Section 4 proposes a quality description metadata format based on the MPEG-7 standard, which allows exchange and usage of analysis results in a flexible way. Section 5 presents visualization and summarization techniques to provide efficient human interaction with quality analysis results. Section 6 provides conclusions and future work.

2. USING RESTORATION ALGORITHMS FOR QUALITY ANALYSIS

2.1. Requirements

The requirements for quality analysis algorithms used in a software based restoration process are as follows. Algorithms should work fully automatic as any human interaction would increase the costs. Analysis can be done off-line before restoration. For seamless integration with software based restoration systems the quality analysis should also be implemented in software. Extensibility and flexibility of a software based implementation is preferable over a hardware based solution. The algorithm's run-time must be low enough to build an analysis system with a manageable number of computers. This implies approximately real-time throughput for a single defect detector. The throughput of multiple defect detectors can be improved by usage of multiple computers in parallel. The algorithm should provide abstracted information about the content. That can be statistical quality measures, e.g. dust or noise level per shot or a listing of certain defect events e.g. missing frame or large dropout events. Only abstract information can be visualized in a compact way, which is a pre-requisite for efficient human inspection of analysis results.

2.2. Algorithms

This section describes how existing impairment detection/restoration algorithms are modified in order to fulfill the requirements stated above.

2.1.1. Dust

Dust is a very frequent defect in historical films. The visible effect in a movie is the appearance of bright or dark spots in single frames. The basic dust detection algorithm is a motion compensated three frame algorithm. It produces a dust mask by observing for each pixel the brightness along the motion trajectory [9]. From a dust mask we can derive the area occluded by this defect as a measure for the dust level of a single frame.

The basis of a shot-based description of dust is the assumption that this artifact is distributed uniformly within the shot or even film material. The measure that represents the dust level of a shot is estimated by the median of all single frame dust areas. The motivation for choosing the median is given in Section 3.

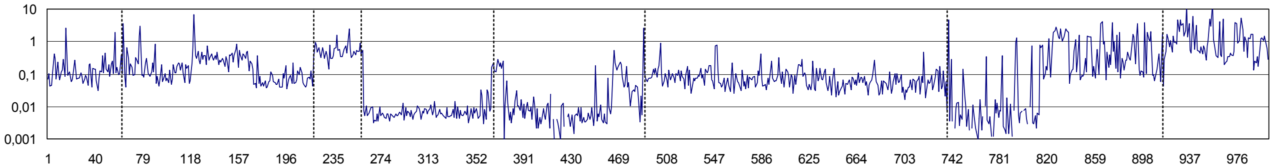


Figure 1: Dust level (percentage of image area occluded by dust) for 1000 frames of the 'Graz' evaluation data. Shot boundaries are indicated by vertical dashed lines.

In order to improve run-time the basic dust detection algorithm is applied only to a reduced number of frames within the shot. Section 3 provides an evaluation on how the quality changes when detection is only applied to a subset of the frames.

2.1.2. Film grain noise

Film grain is an inherent artifact of analog film stock and has significant impact on the effectiveness of digital film processing. Film grain is linked to the physical character of photographic film and is perceived as a random pattern caused by stock dependent local-density variations in an area of uniform density. Our approach for detecting film grain in a single frame considers specific features of film grain noise and aims at detecting regions containing pure film grain and information about its signal dependency [8]. These grain regions provide an indication of the amount of grain, commonly expressed in the peak-signal-to-noise ratio (PSNR) measure.

Film grain characteristics depend on the film stock, the film lighting conditions, the non-linear relation between film exposure and density and on the non-linear operations possibly applied within the digitization process. We assume that these properties are constant in a visual shot, which allows a shot-wise analysis and description of the amount of grain.

Film grain can most clearly be captured in homogeneous areas of the material, resulting in a high PSNR. If a single frame has few homogeneous regions and more textured image content, results are less stable and the PSNR decreases. The larger the PSNR value of a grain region, the less textured image content and the purer film grain is present.

In order to make the film grain level detector as reliable as possible we analyze multiple frames of a shot (every k^{th} frame) and estimate the shot PSNR by the maximum of the single frame PSNR values. The maximum PSNR operation reduces the influence of single frame detector errors due to textured image content in an optimal way. The algorithm presented in [8] has a run-time of about 80ms per frame. In order to reach the real-time requirement for detecting the shot grain level this algorithm can be applied up to every second frame (k=2). To apply the algorithm to a higher number of frames within the shot would not improve the detection reliability as image content does not change significantly between every second frame.

The detection algorithm [8] and assumptions stated above also hold for detecting the electronic noise level per shot.

2.1.3. Flicker

Flicker is a frame-to-frame intensity variation, for example caused by varying exposure of the frames. While there are variants of flicker that occur only locally in the image, global flicker is the more common defect. State of the art flicker detection algorithms (e.g. [10]) calculate an intensity value correction function between the image to be restored and one or more reference images. The challenge is to distinguish between intensity variations due to flicker and others causes, e.g. object motion. Once the correction function is estimated it can be applied to modify the intensity curve of the input image.

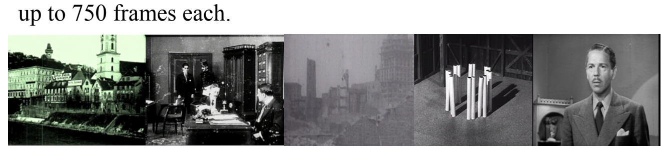
As a result of flicker defect analysis we want to retrieve the average intensity of flicker per visual shot. The use of the following three assumptions about the defect allow for a significant reduction of the processing effort: (a) the flicker is global, (b) it affects the intensity (flicker in a single color channel is rare) and (c) flicker is usually present over a longer time. We thus work on a downscaled luminance image sequence and sample one or more frames per shot that serve as reference images. For k frames before and after the reference frames, we calculate the correction function using an approach similar to that proposed in [10], but using the median of the pixel difference histogram values for stability reasons. Ideally each grey value is mapped to itself, if there is no flicker. We can thus determine the per-frame flicker intensity from the average amplitude of the estimated correction function. The flicker properties per shot are then described by the mean, median and maximum of the per-frame flicker intensities.

3. EVALUATION

For evaluation of the proposed algorithms we have exemplarily selected the dust algorithm. This algorithm has, compared to flicker and grain, the longest single frame detector run-time, therefore for the dust algorithm it is most difficult to reach the throughput requirements stated in Section 2.

For the dust detector we evaluate which run-time improvement and which error is introduced on the estimated dust level when the number of frames analyzed within a shot is reduced. Furthermore this section illustrates the rationale of selecting the median measure for combining single frame dust level values to a dust level per shot.

The evaluation data used are 25 shots from five different movies. Sample poster images are shown in Figure 2. The movies (c), (d) and (e) are publicly available from the Open Video Project [7]. Data is in standard definition resolution; the shots contain 150 up to 750 frames each.



(a) (b) (c) (d) (e)

Figure 2: Evaluation data: (a) 'Graz' (1915) by courtesy of Richard Zumpf, (b) 'Fantomas' (1913) by courtesy of Gaumont, (c) 'Edison Newsreels: San Francisco Earthquake aftermath' (1906), (d) 'Lucky Strike Cigarette Commercial: Square Dance' (1948), (e) 'A Flawed Family Flick - The Boat' (2006).

Figure 1 shows the dust level for each frame of 8 shots, i.e. 1000 frames in total. In Section 2.1.1 we stated that the median of the single frame dust levels best represents the shot-wise measure. As can be seen in Figure 1, in the vast majority of frames within a shot a similar dust level (base line) is detected. For only a small number of frames outliers mainly in the positive direction are detected. The outliers are due to big distortions other than dust, such as short scratches and partially or fully distorted frames. The median represents the base-line dust level well and is less affected by the outliers as the mean. Since the percentage of outliers is usually in the range of 5-20% (depending on the frequency of big distortions), it is valid to compute the amount of dust in the shot i by taking the median of single frame values computed from each k^{th} frame as a representative measure for the shot, in the following denoted by d_{ik}.

The mean relative error E_k for sampling interval k is computed as follows:

$$E_k = \frac{1}{N} \sum_{i=1}^{N} \frac{d_{ik} - d_{i1}}{d_{i1}} 100$$

where d_{i1} is the dust level obtained by computing all frames from shot i. The number of shots is denoted by N.

Figure 3 plots the run-time of the dust detection and the mean relative error E against k sampling intervals, where k runs from 1 to 100.

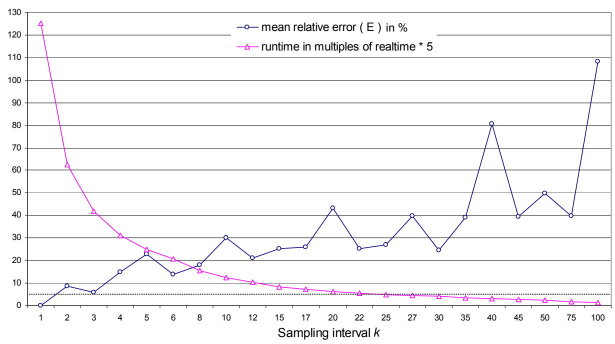


Figure 3: Run-time and mean relative error of dust levels for sampling intervals from 1 to 100. Real-time run-time is indicated by the vertical dashed line (y value of 5).

The dust level estimation error increases with larger sampling intervals. The tradeoff between run-time and error is of great importance. The figure points out that a sampling interval larger than 25 frames is needed to meet the real time requirement for the used dust defect detection algorithm.

Subjective tests have shown that frame to frame and shot to shot dust area variation which are less than 100% are hardly recognizable by humans. For a sampling interval between 25 and 35 frames the mean dust level error introduced by the run-time optimized algorithm is below 43%.

4. DESCRIPTION

As discussed in Section 2, there exist different applications for using the defect and quality analysis results. In order to facilitate interoperability and exchange of defect and quality descriptions between these applications, a standardized way of description must be used.

The description must be able to represent all the results obtained by automatic quality analysis tools as well as additional annotation made by operators. It must both support gaining a quick overview of the overall quality, type and severity of the defects present in the material, as well as describing the detailed measures returned by the tools when applicable. As the defect analysis is targeted towards defect restoration, the description is focused on the appearance of the defects and less on their origin.

4.1 Using MPEG-7 for defect and quality description

MPEG-7 [4] is a standard for the description of multimedia content, including structuring the content as well as describing a number of low-, mid- and high-level features for each of the segments in the structure. Due to the flexibility of the spatial, temporal and spatio-temporal structuring capabilities it is well suited for the description of defects and quality measures that have different temporal and spatial scopes. The defect and quality description is based on the MPEG-7 Detailed Audiovisual Profile (DAVP) [2].

The original version of the standard provides very simple means for describing a quality rating and listing defects present in a segment, but without the capability to specify more in depth information. An amendment to the MPEG-7 audio part [5] defines a more detailed description of audio signal quality, allowing to describe a set of measures per segment as well as a list of error events with different temporal scope and further properties.

4.2. MPEG-7 extension for visual defect and quality description

Based on the existing work in the audio part, we have defined a similar description framework for the visual domain with even more capabilities for describing details of defects. A list of defects and quality measures can be described for each segment. The discrimination between quality measures and defect events described in Section 2 is also reflected in the description: quality measure descriptors contain the statistics for a segment, while defect descriptors describe an occurrence of a defect in more detail.

There is a generic visual descriptor for defects which specifies general properties and references the defect in a classification scheme (cf. Section 4.3). This is the minimum description of a defect, specifying the type of defect and the segment of its occurrence. In addition, specific descriptors for a number of defects and quality measures have been defined, which allow to describe their respective properties. The MPEG-7 extension for defect and quality description is available at [1].

4.3. Impairment classification scheme

The MPEG-7 standard defines a few very small classification schemes for defects. For a detailed specification of the type of defect or quality impairment, a more comprehensive classification scheme is needed. Starting from the BRAVA broadcast archive programme impairments dictionary [6] we have defined a comprehensive impairment classification scheme that provides for hierarchical organization and multilingual description of defects. The main organization criteria of the classification scheme are the visible and audible effects of defects. The impairment classification scheme is also available at [1].

5. VISUALIZATION OF DEFECT AND QUALITY ANALYSIS RESULTS

The visualization of defect and quality analysis results must support the user in quickly getting an overview of the condition of the material. For that purpose, we have implemented the defect and quality summary viewer shown in Figure 4. The tool supports the user in efficiently navigating the content by providing a timeline representation of a number of views. All views are synchronized with the video player. The temporal resolution can be changed so that the user can freely change the level of detail shown. The timeline views show the shot structure of the material, selected representative key frames, stripe images created from the central columns of the images in the sequence and a number of graphs visualizing defects and quality measures. In the screenshot one of the graphs shows the visual activity, which is not a quality measure, but a helpful indicator in the context of restoration. High visual activity indicates either large scale defects (e.g. blotches) or a high amount of motion, which often complicates the restoration process. The other graphs show the shot-wise dust level as the median fraction of the image area covered by dust and grain noise as the image to grain noise ratio.

The temporally condensed overview allows the user to quickly grasp the frequency and strengths of the impairments in the material. From the statistical measures for the individual defects, especially dust and noise level, the needed restoration steps and tools can be planned. Together with the severity of the defects and the user's knowledge about the capabilities of the restoration tools in use the required manual effort and the restoration costs can be estimated. For those defects which are described as defect events, such as big distortions or missing frames, the user can estimate the restoration effort directly from the number of events. Defect event information can also be used for direct examination and restoration of these time intervals in the movie, without having to view the rest of the material.

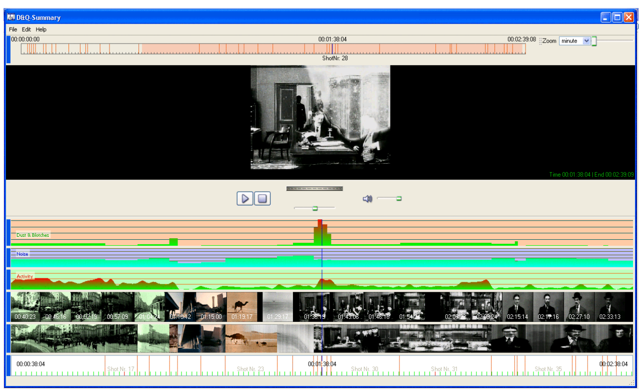


Figure 4: The defect and quality summary viewer.

6. CONCLUSIONS AND FUTURE WORK

Restoration is a new application field for automatic quality analysis tools. Requirements of the software based restoration process are presented.

The real-time throughput requirement enforces the adaptation of existing defect detection algorithms. Adaptations for dust level, grain level and flicker level detection are presented. Evaluation of the dust level detector shows that the mean error of detected dust level is below 43% while the run-time can be reduced by a factor of 35 compared to a dust level detection on all frames of the movie. This fulfills the real-time throughput requirement.

To facilitate interoperability and exchange of quality analysis results a description format for visual defect events and statistical quality measures is proposed. This format extends MPEG-7 in a standard compliant way. Due to the flexibility of the spatial, temporal and spatiotemporal structuring capabilities it can be concluded that MPEG-7 is well suited for the description of visual defects and quality measures.

A defect and quality summary viewer is presented, which supports the user in efficiently exploring and navigating the content by providing a timeline representation of a number of views synchronized with the video player. For several minutes of content the shot structure, key frames, visual activity, dust level and grain level can be investigated at a glance. Restoration cost estimation, planning and manual restoration times can be reduced due to the defect and quality summary visualization.

Future work is targeted to develop and evaluate additional defect and quality detectors as there are image instability, splice and dropout. Furthermore the efficiency improvement in respect to restoration cost estimation, restoration planning and manual restoration effort will be evaluated in a real production environment.

5. ACKNOWLEDGEMENTS

The authors would like to thank Werner Haas, Hannes Fassold and Laurent Joyeux as well as several other colleagues at JOANNEUM RESEARCH, who contributed valuable input to the work. This work has been funded partially under the 6th Framework Programme of the European Union within the IST project "PrestoSpace" (IST FP6-507336, http://www.prestospace.org).

11. REFERENCES

[1] *Audiovisual Defect and Quality Description.* URL: http://mpeg7.joanneum.at.

[2] W. Bailer, P. Schallauer, "The Detailed Audiovisual Profile: Enabling Interoperability between MPEG-7 Based Systems," *Proc. of 12th International Multi-Media Modeling Conference*, Beijing, CN, Jan. 2006.

[3] *Brava final project report.* URL: http://brava.ina.fr/Brava_D12_final_proj_report_b1.pdf, pp. 6-7.

[4] ISO/IEC, Multimedia Content Description Interface, ISO/IEC 15938:2001.

[5] ISO/IEC, Multimedia Content Description Interface, Part 4: Audio, ISO/IEC 15938-4:2002/Amd 1:2004.

[6] *The Brava broadcast archive programme impairments dictionary*. URL: http://brava.ina.fr/brava_public_impairments_list.en.html.

[7] The Open Video Project. URL: http://www.open-video.org/

[8] P. Schallauer, R. Mörzinger, "Rapid and Reliable Detection of Film Grain Noise," *IEEE Intl. Conf. on Image Processing*, Atlanta, USA, Oct. 2006.

[9] P. Schallauer, A. Pinz and W. Haas, "Automatic restoration algorithms for 35mm film," *J. Computer Vision Res.*, 1(3):59-85, 1999.

[10] T. Vlachos, "Flicker correction for archived film sequences using a non-linear model", *IEEE Trans. on Circ. and Sys. for Video Tech.*, 14(4):508-516, April 2004.