Automatic Content Based Video Quality Analysis for Media Production and Delivery Processes

Peter Schallauer, Hannes Fassold, Martin Winter, Werner Bailer, Georg Thallinger, Werner Haas

JOANNEUM RESEARCH Forschungsgesellschaft mbH
Institute of Information Systems
Steyrergasse 17, 8010 Graz, Austria

Automatic quality control for audiovisual media is an important tool in several steps of the media production, delivery and archiving processes. Today, mainly technical properties of the material are checked, e.g. stream compliance, playtime, aspect ratio, and resolution or MXF compliance. Only some content properties can be checked automatically, e.g. blocking or luma/chroma violation. Other relevant content properties and impairments like noise level, sharpness, large dropouts, flickering or instability are checked by manually exploring the audiovisual content. In this work we focus on challenges and recent results in automatic content based visual quality analysis of video. We first give an overview on which visual impairments are relevant in which stages of the media production, archiving and delivery process. A set of requirements for impairment detection algorithms, tools and systems is presented. We show how impairment detection algorithms need to be designed in order to meet these requirements. Furthermore we show our recent algorithmic research results for two content based impairment detectors (freeze frame and video breakup detection). In order to facilitate interoperability and exchange of impairment metadata between different tools and systems, a standardized way of description is needed. We give an overview on our framework proposed for the description of visual impairments based on MPEG-7. In order to enable efficient human interaction with quality analysis results we present the "Quality Summary Viewer" application which allows a user to quickly grasp the frequency and strengths of visual impairments in the content.

1 Introduction

Automatic quality control for audiovisual media is an important tool in several steps of the media production, delivery and archiving process. Broadcasters are checking quality at ingest, after editing and before play-out to various delivery services. Archives are checking for content integrity at archive ingest and delivery. Content providers are checking their content for correct encoding and conformance to the required format standard before dispatching to customers. These use cases have in common that mainly technical properties of the material are checked, e.g. stream compliance, playtime, aspect ratio, resolution or MXF compliance. Today only some content properties can be checked automatically, e.g. blocking or luma/chroma violation. Other relevant content properties and impairments like noise level, sharpness, large dropouts, flickering or instability are usually checked by humans manually exploring the audiovisual content. Figure 1 indicates which video and film impairments may appear within the media production, delivery and archiving processes.

In this work we focus on challenges and recent results in automatic content based visual quality analysis of video. For this purpose we have organized the paper as follows. Section 2 presents a set of requirements for impairment detection algorithms, tools and systems for software and file based environments. It outlines general design criteria for such algorithms in order to meet these requirements. In addition, we present our recent algorithmic research results for the content based impairment detectors freeze frame and video breakup. In section 3 we give an overview on our framework proposed for the description of visual impairments. In section 4 we present the Quality Summary Viewer application developed for efficient visual impairment visualization and exploration.

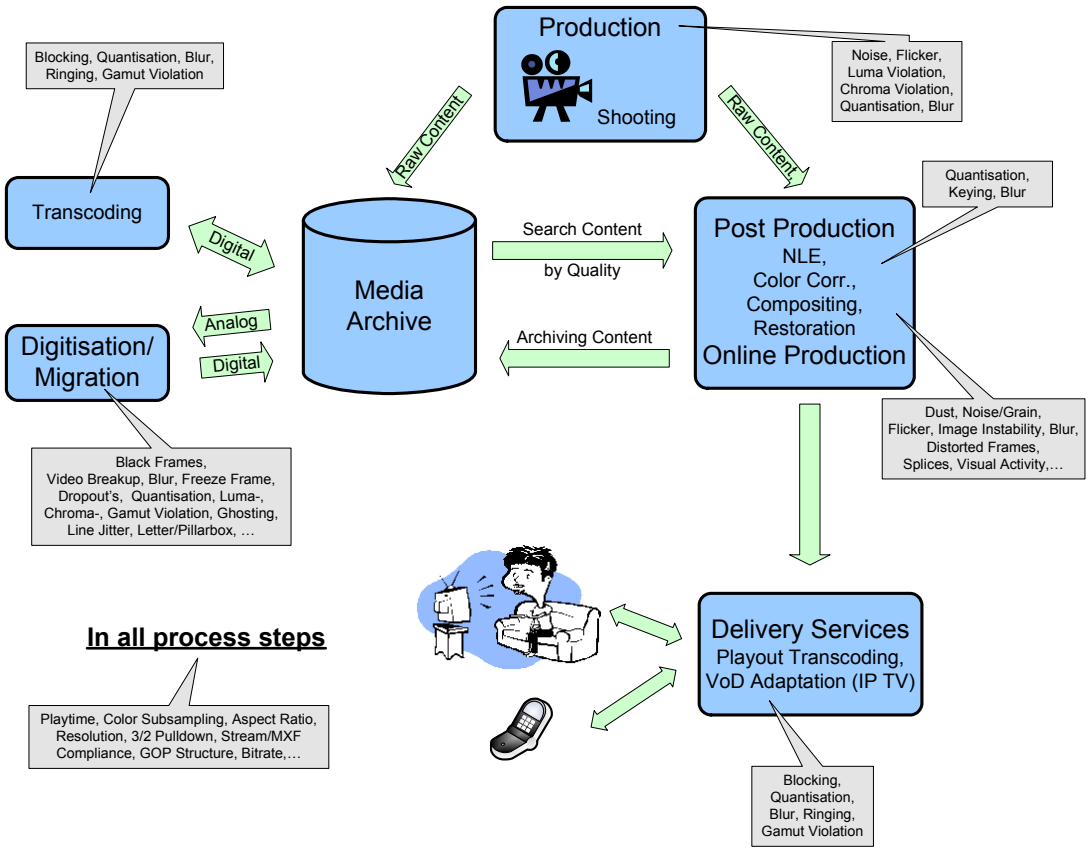


Figure 1: Video and movie impairments appearing during the media archive-, production-, and delivery processes.

2 Impairment Detection Algorithms

2.1 Application Requirements

In order to design algorithms for the purpose of content based video impairment analysis it is important to understand the application requirements for those algorithms. In the following we present a set of requirements for impairment detection algorithms, tools and systems for software and file based environments.

The algorithms should work fully automatic or at least semi-automatically as any human interaction increases the costs. Quality/impairment analysis can be done off-line before judging and using the impairment information. For seamless integration with software based application environments quality analysis should also be implemented as much as possible 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 at least real-time throughput for a single impairment detector. The throughput of multiple impairment detectors can be scaled by adding more processing nodes. 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. freeze 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. For each of the algorithms the aim is to report analysis results correlating with human perception of the impairment. Thus the consideration of content properties and other defects is important as they might significantly influence the perception of the impairment strength. This means, that the final assessment of each of the impairments cannot be done independently but needs to be done on higher level, taking the interdependencies between the different impairments into account.

2.2 Algorithms Design Criteria

In order to meet the application requirements stated in section 2.1, certain general design criteria for impairment detection algorithms can be applied.

One major issue for software based implementations are strategies to meet the real-time speed requirement. Certain impairments appear spatially global, e.g. global flicker. For this type of impairments speedup can be achieved by applying a spatial sub-sampling strategy within the algorithm. Other impairments show temporally constant properties over time. A typical example therefore is e.g. the noise or dust level which usually changes only slightly within a shot. For this type of impairments algorithmic speedup can be achieved by temporally sub-sampling the impairment detection. In [SBMFT07] this strategy has been applied in order to efficiently detect the amount of dust spots within movie shots. For impairments where it is not possible to apply spatially or temporally sub-sampling within the algorithm one option for improving speed is to implement computationally intensive parts of the algorithm within hardware accelerated environments. Beside FPGA based technologies graphics processing units (GPU's) gain special attention due to their cost efficiency. A prerequisite for GPU based acceleration is that the algorithm must be massively parallelizable. In this case a speedup factors of five to hundred compared to a purely software based implementation is achievable.

Interdependencies between different impairments pose a special challenge for implementing an automatic system with acceptable detection performance. For the development of the individual impairment detection algorithms this requires robustness against influences from other impairments present in part of the material. This also requires adaptation of the algorithm to other impairments in a certain segment of the content.

2.3 Freeze Frame Detector

Freeze frames (or fields) occur, when due to various reasons no valid data for the current frame (field) can be retrieved. In this case, most video devices deliver the previous frame (or field) instead. The effect is that consecutive frames will have equal image content. The difficulty in freeze frame detection is that static video sections occur regularly in video content and should not be regarded as freeze frame impairment. Examples are captions, full screen logos or short static section within a low-motion scene (e.g. actor stands still for a moment). In order to handle this, we developed a combined algorithm which is composed of a basic detector delivering a set of *potential* freeze frame segments, followed by a high-level analysis which classifies these potential freeze frame segments as normal content or as an actual freeze frame impairment.

The *freeze frame basic detector* identifies segments in the video stream with zero within-segment content change. For this purpose, it uses two different activity measures for measuring the amount content change between two images: The *visual activity* $A(I_1, I_2)$ is calculated statistically from the difference image. Due to a special design it it is robust to noise and slight brightness changes, which are not perceived as content change by a human. As a complementary measure the *block motion activity* $B(I_1, I_2)$ is used, which calculates the average block motion between the two images. It is useful for scenes with low-contrast, slow-moving content (e.g. moving clouds) where the visual activity measure reports very low values. Due to real-time runtime requirements the calculation of the block motion is done using integral projections [KP92]. Video segments with low within-segment visual and block motion activity are considered by the basic detector as *potential* freeze frame segments (PFFS). Some post-processing is done to merge temporally neighboring PFFS which have been split up by accident and to discard PFFS within a scene with low motion. A detailed description of the *freeze frame basic detector* can be found in [SFWB09].

The *high-level analysis* takes as input the potential freeze frame segments (PFFS) provided by the basic detector. For each PFFS, it calculates a set of features. Some MPEG-7 image descriptors are calculated for the frames within the PFFS, e.g. the color dominance and the color layout descriptor. Additionally, it calculates some temporal statistics (e.g. PFFS length, ratio of shot length to PFFS length, location of PFFS within shot). Based on this feature set and on a given ground-truth annotation for FFS impairments, a support vector machine (SVM) is trained which then classifies a PFFS as actual freeze frame impairment or normal static content.

We evaluated the combined algorithm on a large video database, containing 7.2 hours of material with very different content characteristics (e.g. high-speed sport scenes, slow pans in weather surveillance camera transmissions, static film captions) for which 45 freeze frame impairments were manually annotated. It is important to note, that the material contains also other impairments like noise, flicker, blocking etc. For this data set, the combined algorithm achieves a recall of 72% and a precision of 94%. Note that the basic detector alone has a slightly higher recall of 84%, but a precision value of only 11% due to many false positives in static video sections (e.g. captions). So the high-level analysis is absolutely essential to achieve low false-positive rates.

2.4 Video Breakup Detector

Video breakup is an umbrella term for a set of strong visual distortions typically caused by tape transportation problems on analog video material or (partially) broken digital video streams. The appearances of this impairment vary substantially and it is difficult to identify common patterns for reliable defect detection. Some typical examples for video breakup events are depicted in Figure 2.

The basic idea behind our proposed algorithm is to distinguish all kinds of normal object motion within the content from abrupt content changes induced by video breakup. Video breakup distortions significantly change their location and appearance from frame to frame. As mentioned in section 2.2 the impairment detection algorithms are subject to several design criteria, especially runtime constraints prohibit the usage of complex and therefore time consuming algorithms. In its original form, optical flow calculation has been a very time-consuming task and was therefore rarely used for applications with real-time requirements. Nevertheless, recent research takes advantage from the highly parallelized architecture of graphic processors (GPU) thus providing real-time capabilities for optical flow calculation even on full standard definition resolution videos [ZPB07].

In particular, our proposed algorithm works as follows: At first, we estimate the motion between two consecutive frames I_n and I_{n+1} and use it to warp the image I_n to the motion compensated image $I_{n+1,mc}$. Simply calculating the pixel-wise difference between the estimated images $I_{n+1,mc}$ and I_{n+1} results in a difference image (D) with significant responses on video break-up locations. In a consecutive step we cumulate the difference values in each row of D and obtain a „stripe-difference-image“ representation as e.g. shown in Figure 3: (b) and (d). While regular motion causes only slight changes from column to column, video breakups as e.g. shown in Figure 3: (c) and (d) induce heavy disturbances. Hence, the distance measure from [WNK03] allows us to estimate the probability of video breakup presence for each frame of the video.

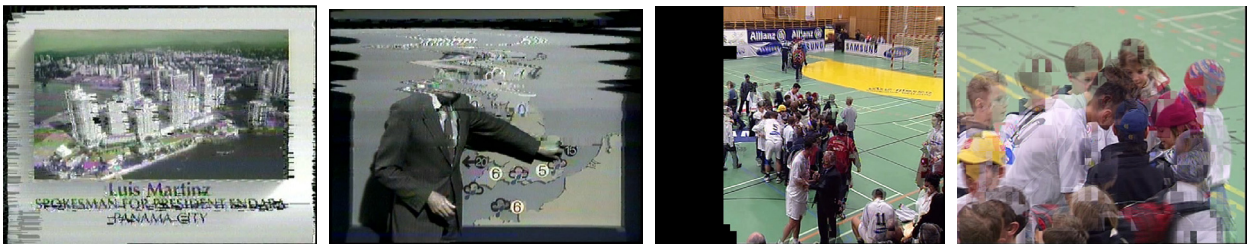
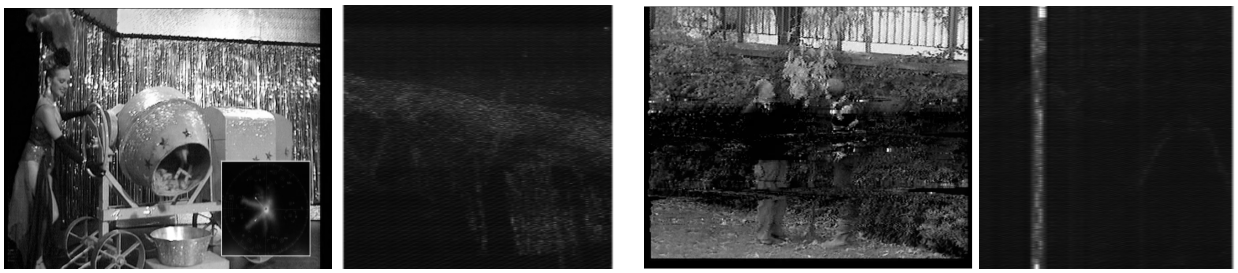


Figure 2: Some typical examples for video breakup impairments caused by analog and digital sources.



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

Figure 3: Some examples for „stripe-difference-images“ (b,d) for a scene containing regular motion (a,b) and video breakup (c,d).

As we found, that the basic video breakup measure proposed occasionally fails in scenes with heavy, suddenly appearing illumination changes (e.g. caused by flash lights or other strong illumination changes), we introduce a second measure based on the ratio (R) between vertical and horizontal edges on the difference image directly. In particular we estimate the averaged change of R within a certain time interval (several frames) to indicate a video breakup event. The basic assumption behind is, that the amount of horizontal and vertical edges changes equally for extensive illumination changes.

To evaluate the performance of our algorithm we use a compilation of several videos with various, challenging contents (e.g. scenes with extremely fast motion, heavy luminance changes, noise, etc.) and a substantial amount of video impairments. For an objective judgment, the video impairments have been annotated by experts from a local video producing company.

For the segments containing video breakups caused by tape transportation problems on analog video materials we obtain a recall/precision rate of 89% and 80% respectively. Including the video breakups from partially broken digital video streams the precision degrades to 53%. The main reason therefore is the increased existence of blocking artifacts, which is actually not covered by our detector, but we are confident of solving this in the near future. The main limitation of our proposed algorithm is given by videos containing extremely fast movements. In such situations, the optical flow calculation fails, and the caused responses in the difference image (D) feign video breakup events. Anyway it is possible to be aware of such events by extremely lowering the algorithms' confidence level in high motion sections detected by inspection of elongated time intervals.

3 Interoperable Impairment Description

3.1 Motivation and Requirements

Standardized impairment description of audiovisual media is a pre-requisite for system interoperability between content digitization, documentation, management, restoration, production and delivery systems. The impairment description shall allow getting an overview of the condition of the audiovisual material. It shall thus be a compact description and contain details only if absolutely necessary. The description shall not include intermediate results of specific restoration algorithms, configurations of analysis or restoration and a history of applied restoration steps. The description is mainly produced by automatic tools, and it shall also be possible to process the description automatically. Therefore,

- the time point or range for which a description is valid must be specified,
- quality has to be quantified numerically or by sets of defined terms,
- defects need to be unambiguously identifiable, and
- optionally, properties of defects may be further described numerically or by sets of defined terms.

As the descriptions support the user in getting a quick overview of the materials condition, they shall be defined in a way that they are easy to visualize. Especially quality measures and defect descriptors that represent a larger time range shall allow condensed visualization over time. Quantitative descriptions of impairments shall correspond to the perceived severity of the defect.

3.2 State of the Art

While most multimedia metadata standards allow the description of technical signal parameters of the audio and visual content, capabilities for describing impairments are rare. The SMPTE Metadata Dictionary [RP210] contains elements for describing overall assessments of the technical quality, some video test parameters, the audio and visual signal to noise ratio and quality events and parameters of audio data as defined by the Broadcast Wave Format (BWF [BWF]).

MPEG-7 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. In MPEG-7 already some impairment descriptors and description schemes have been standardized. The *MediaQuality* descriptor [MPEG7-5] is an element in the *MediaProfileDS* and thus applicable to all media types. It contains (i) quality rating, expressed as a floating point value, (ii) a rating source and a reference to the rating information and (iii) a list of perceptible defects, discriminated into visual and audio defects. Each defect is a reference to a term in a classification scheme. It is not possible to describe the defect in more detail or its exact (spatio-) temporal location. The *AudioSignalQualityDS* has been introduced in amendment 1 to part 4 [MPEG7-4A1]. It can be added to each audio segment and contains the following elements: balance, noise level, DC offset, cross channel correlation, delay, a list of error events. Each of these error events is described by the error class (a reference to a term in a classification scheme), time stamp and channel number, detection method (manual, automatic), relevance, status and optional text annotation. Classification schemes are MPEG-7 description schemes for defining hierarchies of controlled vocabulary. The following classification schemes exist for the description of impairment information: *AudioDefectsCS* and *VisualDefectsCS* [MPEG7-5] and *ErrorClassCS* [MPEG7-4A1].

3.3 MPEG-7 Extension for Visual Defect and Quality Description

Based on an analysis of the state of the art and the requirements defined above it becomes clear that MPEG-7 is a suitable standard to serve as a basis for the description of visual impairments. MPEG-7 allows to structure descriptions on different levels of granularity and already offers some tools for quality description, especially in the audio domain. We have thus chosen the following steps for extending MPEG-7:

- Define a description scheme for visual impairments, similar to that for audio quality and defects defined in [MPEG7-4A1].
- In addition, allow the extension by detailed descriptors for specific quality measures and defect descriptors.
- Define these specific descriptors for some common quality measures and defects.
- Extend the existing MPEG-7 controlled vocabularies (classification schemes) for both visual and audio impairments.

The extension is based on the MPEG-7 Detailed Audiovisual Profile [BS06] which has been proposed for detailed description of audiovisual content in production and archiving. The MPEG-7 extension for defect and quality description is available at [JRSDQ].

There is a generic visual descriptor for defects which specifies general properties and references in a classification scheme. This is the minimum description of a defect, specifying its type 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 following specific descriptors have been defined:

Dropout/partial frame damage describes a dropout or other damage affecting a part of a single or a few frames. The lines or region and the channels affected can be described.

Full frame damage, freeze frame, black frame or lost frame describes the loss of a frame, the damage of a frame or the replacement of a lost frame by a previous one (freeze frame).

Video breakup describes salient visual distortion of one or several subsequent frames, the area affected can be annotated.

Line scratches describe vertical scratches on film material occurring in arbitrary temporal segments from a few to hundreds of frames and their properties (horizontal position, width, and negative/positive).

Number of line scratches describes the number of scratches in a shot.

Dust/dirt level describes the level of dust or dirt spots level per shot or part of a shot. The average number and size of spots and the average intensity can be described.

Noise/grain level describes the noise or film grain level of a shot, using the properties PSNR, spatial frequency and brightness dependency.

Flicker level describes the level of temporal intensity variations (flicker) in a shot, described by the average flicker intensity, its frequency distribution and the local variation.

Image instability describes the geometric position instability of the image by the average/maximum horizontal/vertical displacement.

Loss of resolution describes the loss of spatial resolution of image (e.g. due to up-conversion) in a shot by characteristic scale of the edges.

Blocking level describes amount of blocking artifacts level from lossy DCT based encoding per shot or part of a shot.

Dropout level describes the number and area of dropouts per shot or part of a shot.

Channel misalignment describes asynchronous color channels in a shot or the whole material.

Color range defect describes high-/low contrast in one or more channels or saturation/clipping in a sequence of shots or the whole material by specifying the fraction of the intensity range used and/or the area affected by saturation.

3.4 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 [BRAVA] 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 top level elements of the impairment classification scheme are shown in Figure 4 and it is also available at [JRSDQ].

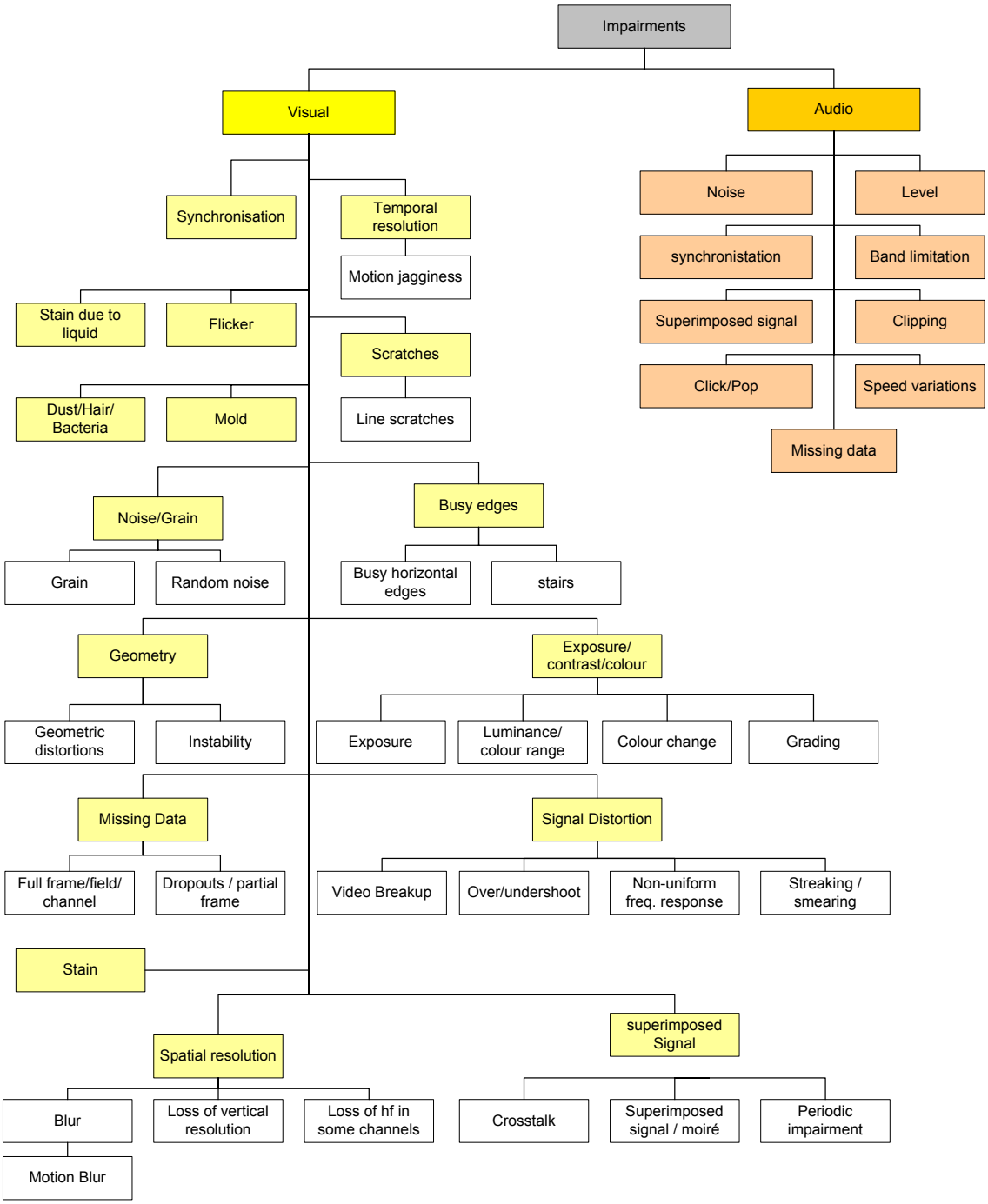


Figure 4: Top-level classes of ImpairmentCS.

4 Result Presentation and User Interaction

The visualization and exploration of impairment analysis results must support the user in quickly getting an overview of the condition of the material. For that purpose, we have implemented the Quality Summary Viewer application shown in Figure 5. 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 a restoration application. 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.

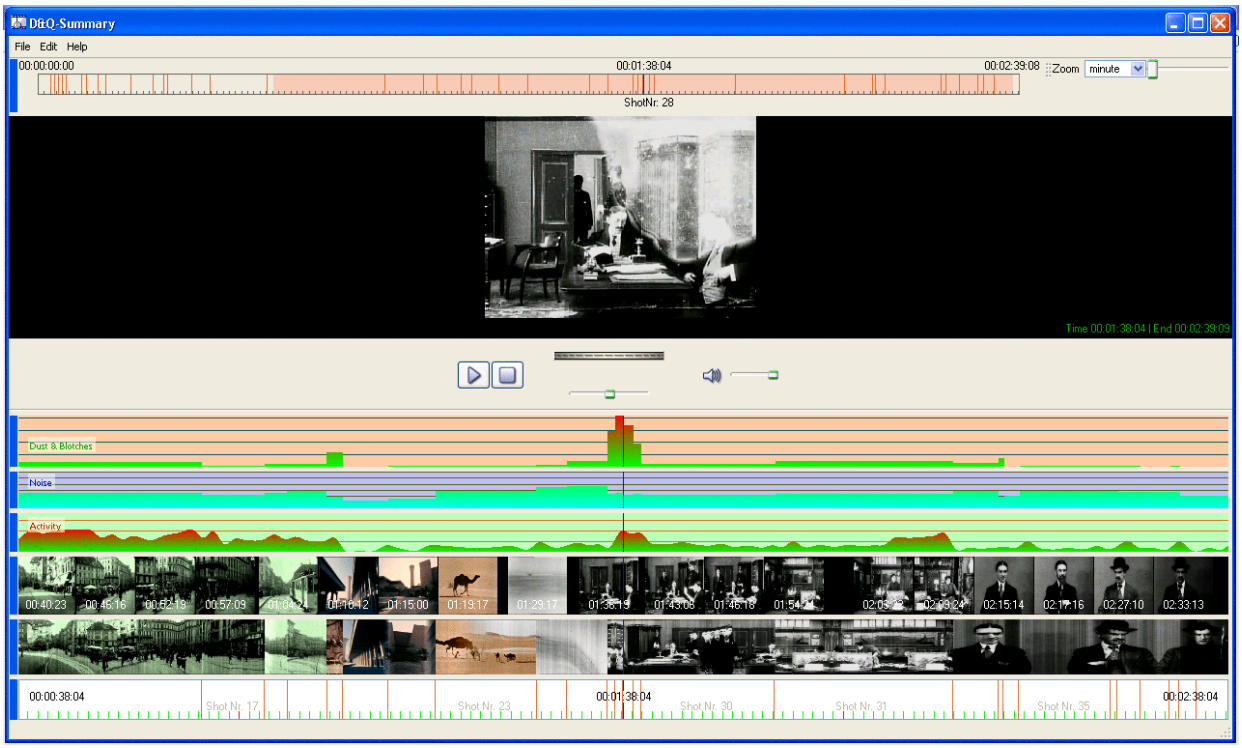


Figure 5: Screenshot from our Quality Summary Viewer.

5 Conclusions

Today content based quality analysis is applied only for some impairments appearing in the media production, delivery and archiving processes. One reason is that for content based quality analysis algorithms very challenging application requirements need to be fulfilled. Algorithms should run fully automatically in real-time, software implementations are preferable, algorithm should provide abstracted information in form of quality measures and defect event listings and analysis results should correlate with human perception also taking into account interdependencies between the different impairments.

Spatial and temporal sub-sampling within the algorithm is an efficient strategy for achieving the real-time requirement. In cases where this cannot be applied GPU implementation is an attractive alternative. Impairment detection algorithms need to be made robustness against influences from other impairments which may be present.

The proposed freeze frame impairment detector achieves a recall of 72% and a precision of 94% due the usage of a combined algorithm which is composed of a basic detector and a second learning high-level analysis step. The second impairment detection algorithm presented focuses on video breakups and shows promising results especially for breakups originating from analog video by achieving a recall rate of 89% and a precision of about 80%.

To facilitate interoperability and exchange 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, and several impairment measures can be investigated at a glance.

6 Acknowledgements

The authors would like to thank Hermann Fürntratt, Albert Hofmann, Marcin Rosner as well as several other colleagues at JOANNEUM RESEARCH, who contributed valuable input to the work. This work has been funded partially under the 7th Framework Programme of the European Union within the IST project "PrestoPRIME" (IST FP7 231161) and under the FIT-IT Programme of the Austrian Federal Ministry for Transport, Innovation and Technology within the project “vdQA”.

7 References

[RP210] Metadata Dictionary Registry of Metadata Element Descriptions. SMPTE RP210.8, 2004.

[BWF] BWF – a format for audio data files in broadcasting EBU Tech 3285, 2001.

[MPEG7-4A1] ISO/IEC, Information Technology – Multimedia Content Description Interface, Part 4: Audio, ISO/IEC 15938-4:2002/Amd 1:2004.

[MPEG7-5] ISO/IEC, Information Technology – Multimedia Content Description Interface, Part 5: Multimedia Description Schemes, ISO/IEC 15938-5:2001.

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

[BS06] Werner Bailer and Peter Schallauer, “The Detailed Audiovisual Profile: Enabling Interoperability between MPEG-7 Based Systems,” Proc. of 12th International Multi-Media Modelling Conference, Beijing, CN, Jan. 2006.

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

[KP92] J. Kim and R. Park, “A Fast Feature-Based Block Matching algorithm using Integral Projections”, IEEE Journal on selected Areas in Communications, Vol. 10, No. 5, 1992.

[SBMFT07] Peter Schallauer, Werner Bailer, Roland Mörzinger, Hermann Fürntratt, Georg Thallinger, “Automatic Quality Analysis For Film And Video Restoration,” Proc. IEEE International Conference on Image Processing, 2007.

[SFWB09] Peter Schallauer, Hannes Fassold, Martin Winter, Werner Bailer, “Automatic freeze frame Detection for Video Preservation,” Proc. IEEE International Conference on Image Processing, 2009.

[WNK03] Dietrich Van der Weken, Mike Nachtegael, and Etienne Kerre, “Using Similarity Measures for Histogram Comparison,” Proc. 10th International Fuzzy Systems Association World Congress,, pp.1-9, 2003.

[ZPB07] Christopher Zach, ThomasPock, Horst Bischof, “A Duality Based Approach for Realtime TV-L1 Optical Flow”, Proc. 29th DAGM Symposium on Pattern Recognition, pp. 214-223, 2007.