International Journal For Technological Research In Engineering
Volume 6, Issue 9, May-2019 *ISSN (Online): 2347 - 4718*

APPROACHES FOR QUALITY-AWARE VIDEO CONTENT ANALYTICS

Mr. Manu Y M[1], Harshitha K S[2], Lakshmi R[3], Navya A S[4], Navya A U[5]
[1]Assistant professor, [2,3,4,5]UG Student Department of Computer Science and Engineering
BGS Institute of Technology, B G Nagar, Mandya-571448

ABSTRACT: Recent research in video analytics promises the capability to automatically detect and extract information from video. Potential tasks include object andpedestrian detection, object and face recognition, motion detection, object tracking, as well as background subtraction and activity recognition. However, in many instances, the quality of the video from which information is to be extracted is notvery high. This may be because of system constraints (like a bandwidth constraint or VHS recorder), environmental conditions (fog or low light), or a poor camera (wobbly/moving camera, limited FOV, or just a low-quality lens). In this paper, we provide an overview of research on designing video analytics systems that use potentially low quality data. We consider a variety of analytics tasks, and present five categories of approaches to create quality- aware analytics: quantify the impact, predict the impact, create an analytics-aware encoder, enhance the input before analytics, and modify the analyticsalgorithms.
Index Terms: Video Analytics, Video Quality

I. INTRODUCTION

The goal of video content analytics is to use a machine, not a person, to extract information from video. The extracted information can answer questions of who, what, where, when, and "how well". Video analytics can be retrospective (looking back at the past for forensic evidence), real-time (operating "in the now" and generating alerts), predictive (identifying what is likely to happening the future, perhaps because of some identified anomaly), or prescriptive (deciding what action should be taken based on the data). The ability of automated methods to extract information from video has been increasing dramatically in recent year, in part due to recent successes of training deep neural networks using extensive datasets. In many cases, the systems are trained on clean data: not just data that have been trimmed, cropped, painstakingly labeled, but also data that are mostly free from the types of degradations that pervade real systems. For example, images from the ImageNet classification dataset typically contain one relatively-centered clearly-focused object that fills most of the image frame. When a system is deployed in real-world scenarios, performance can suffer. Many quality degradations occur at the moment of image or video capture (for example, lighting or illumination, or low-light noise or motion blur), while others are "self-indicted" because of constraints within the system (for example transcoding, down sampling, or frame dropping). Low-quality inputs degrade analytics performance. To study the confluence of video quality and video analytics, it is necessary to consider at least 3 distinct topics: the degradations that can be present, the types of problems that video analytics systems can be designed to address, and the methodologies of solutions for those video analytics systems. Each methodology might have its own vulnerabilities, and hence require its own remediation method. In this paper, we begin in Sections 2 and 3 with a short summary of degradations and video analytics, respectively, and then in Section 4 describe five categories of approaches that have been considered in the literature to overcome these challenges. Due to space restrictions, we choose representative referencesonly.

II. VIDEO ANALYTICS OVERVIEW

A typical processing pipeline for a video analytics system includes capture, compression, transmission, analysis, alerting, and storage. Our focus in this paper is to consider the implications of the degradations created during capture, compression, and transmission on the analytics tasks. Three core tasks are recognition, localization, and detection, each of which can be applied to a variety of objects, events, or activities. Specific objects of interest include people, faces, text, vehicles, and license plates, while example actions of interest may include gestures, slip and fall, or leaving a bag. Additional tasks include object, image, semantic, and instance segmentation, scene understanding, 3D-scene reconstruction, summarization, tracking, and generic anomalous event detection. In many cases, there are common computational steps that can be shared across a variety of analytics tasks. Foreground/background segmentation may be useful for object tracking and object detection, while image registration, object classification, tracking, and action recognition may all share the same key point extraction step. In addition, the result of a first task may determine whether a second task is performed or not. For example, if there has been nothing moving in the scene, there is little reason to perform pedestrian detection. Three basic types of features are typically extracted from video to perform video analytics tasks: hand-crafted features, kernel-based descriptors, and features learned using deep convolution neural networks. It is likely that each feature will have its own robustness to impairments in the source video.

III. VIDEOQUALITYDEGRADATIONS

Video degradations are commonplace in deployed systems. Some degradations may happen at the moment of capture, i.e., in the camera, while others may be "self-inflicted", introduced prior to the analysis stage because of constraints within the system. For example, transcoding may happen in the network if bandwidth is limited and there is adaptive

bitrate video streaming. After all, it is often considered better to obtain a low-rate version of the video than no video at all. The quality of video captured by the camera depends on the spatial and temporal resolution, any lens distortions (like fish-eye), rolling shutter, and blur. A poorly placed camera might impair the field of view of the object or action to be identified. In addition, camera motion may degrade visual quality, but may also provide valuable information for analytics. For example, a camera mounted on the body or the head may provide useful information to identify human behavior. Finally, the camera's viewpoint and location may create obstructions, and an elevated camera may alter the perspective on objects within the field of view. Environmental viewing conditions can also impair the video during capture. In particular, lighting and illumination can create glare, reflections, and under-exposed video. Low light conditions can also introduce noise. Rain can cover the camera lens and create distortions, clouds may create time vary in gillumination issues, while fog and snow may decrease quality by reducing contrast. Finally, video compression and transmission- induced packet lossesor RF interference can all decrease video quality between the camera and the analysis stage. Video compression depends on the available bandwidth for communication, and it may be un avoidable when it is imposed by system constraints. Each of thecamera, compression, and transmission may introduce blur or noise. Each analytics task may have its own specific quality requirements. Recognition may require sharper, higher-quality inputs either than detection or than analytics on larger objects using simple motion or flow. Object recognition performance depends on resolution, lighting and illumination, blur, occlusion angle, field of view. The performance of tracking depends on object speed, shape, and deformations. Tracking may need higher frame rate than classification, but classification may need a higher spatial resolution. Moreover, the frame-rate during tracking should be adequate to capture an object's speed and motion. As mentioned above, in some cases the degraded quality may actually be informative for the analytics system. For example, camera motion may either be considered to be a degradation (when it detracts from the ability to identify an object), or actually be the information to be extracted (when the camera is mounted on anobject).

IV. QUALITY-AWAREVIDEO ANALYTICS

The prior work on the topic of video analytics using degraded quality images or videos can be subdivided into 5 basic categories.

- Quantify theimpact.
- Predict theimpact.
- Create an analytics-aware encoder for video orfeatures.
- Enhance the input beforeanalytics.
- Modify the analytics algorithm. We will consider each of these topicsbelow.

4.1. Quantify the impact

Quantifying the impact of a degradation is straight forward: apply the analytics algorithm to a variety of degraded inputs, at increasing levels of severity, for each type of degradation of interest, andcharacterizethe resulting accuracy. The impact of compression on face detection, tracking, pedestrian detection, and activity recognition have been studied in and considers the impact of object size, occlusion, and aspect ratio. It is unnecessary to distribute databases containing synthetically- introduced distortions. Instead, a methodology that repeatably introduces synthetic distortions at the requested level is more flexible. The result of this type of study can allow a system designer to understand the required video quality to achieve a given prediction accuracy, and to determine what resources are necessary to achieve that video quality. For example, it may be possible to characterize the minimum necessary bandwidth in aggregate, and then design the system with lower resource requirements. However, if insufficient resources are available, then it allows performance expectations to be adjusted based on system constraints. In addition, this knowledge can be used to dynamically allocate resources in a camera network. However, it is important to note that the impact of quality on accuracy may be content dependent. It was shown in that to obtain sufficient accuracy for activity recognition, a different minimum quality may be required for each activity. Also, the frame- rate required for accurate object tracking was shown to depend on image content. Predict the impact A variety of approaches have been explored, for predicting the impact of quality degradation on the accuracy of a video analytics algorithm. One basic approach is to consider this prediction as a quality estimation, where quality to be estimated is the analytics accuracy. This differs from the usual no- reference quality estimation, in that the goal is not to characterize how humans will perceive the video. This prediction can happen at the edge of the network, prior to compression, or near the video analytics engine. A quality model to compare the expected accuracy of three moving object detection algorithms before applying compression was proposed. In, a quality model to predict object tracking performance of a specific tracker was designed, when the frame rate and spatial resolution varied. Pedestrian detection was considered, where the predicted accuracy was proposed to be used in two ways: first at the encoder to choose the level of compression that lowers the bit- rate most without sacrificing detection accuracy, and second at the decoder to choose the lightest-weight method (computationally) that will still satisfy the desired accuracy. Pedestrian detection was also considered, with the goal to predict at the decoder, for a given degraded input video, which of several quality-aware analytics algorithmswill provide the best performance. A similar idea was applied for activity recognition, which predicts the level of compression necessary at the encoder for each input video, with the goal of achieving as much compression as possible for each individual video, without affecting the system's ability to recognizeactivities.

4.3. Create an analytics-aware encoder Creating ananalytics-aware encoder is distinct from the above strategies that choose the bandwidth at which to compress the video. Instead, the goal here is to modify the video encoding algorithm to provide the best-for-analytics bitstream at a

given bandwidth, or to design an encoder that operates directly on features already extracted during video analytics. Algorithms that modify the video encoding parameters, to create a standards-compliant bitstream knowing that it will be used solely for analytics, were designed. This allows the encoding parameters to be specifically chosen to minimize the impact of compression on analytics. In the Analyze-Then-Compress strategy (ATC), features are extracted prior to video compression, and a feature specific encoder is created. This avoids the problem of extracting useful features from degraded compressed video, but requires sufficient processing power at the network edge for feature computation. This approach was used for object recognition, tracking, and a hybridapproach that encodes both features and video was proposed for pedestrian detection.Enhance the input prior to analytics

Another straightforward approach is to enhance the video input prior to running the analytics algorithm. Demosaicing, white balancing, and denoising are all applicable here, as is dehazing or contrast enhancement. The ability of deblurring, image interpolation, and single- image super resolution algorithms to improve object recognition is considered. Modify the analytics algorithm A wide variety of directions have been explored to create quality-aware video analytics algorithms. Each can be applied to individual or multiple degradations, and can be generic for all level so fquality or targeted to a specific quality range. Each stage of the analytics algorithm can be modified to account for lower-quality inputs. Blur-robust descriptors for face recognition are designed, andmodifies gradient descriptors to accommodate the distortion caused by a wide-angle lens. Other approaches include domain adaptation and transfer learning, using data augmentation during training to include additional synthetically created distortions, or simply training the system on low-quality data. In addition, performance of a system that will use only low- resolution data can be significantly improved by incorporating highresolution inputs during training. There are ample opportunities to integrate concepts of No-Reference Quality Estimation algorithmsinto various video analytics algorithms. For example, quality-based features were shown to significantly improve performance of a face detection algorithm given low-quality inputs. A systems-based approach was designed, where a quality- estimation algorithm was created to select among a collection of pedestrian detection modules, each of which had been optimized to work as well as possible on some range of quality.

V. CONCLUSIONS

We reviewed approaches to design video analytics systems, for videos which have degraded quality. Due to the wide range of analytics algorithms, potential degradations, and design approaches, there are many opportunities for continued improvement.

