Wireless Networking for Automated Live Video Broadcasting: System Architecture and Research Challenges

Domenico Giustiniano, Vladimir Vukadinovic, Stefan Mangold
Disney Research Zurich, Switzerland
Email: {domenico|vvuk|stefan}@disneyresearch.com

***Abstract*—The majority of sport events are of interest only to a relatively small group of viewers, such as college and regional league competitions. The broadcasting of these sport events is often not economically viable because of the required installation/cabling and the need for an on-site crew for content production. An automated or semi-automated broadcast system with multiple cameras would potentially enable a significant cost reduction. This contribution analyzes to what extent wireless technology simplifies the deployment and reduce the cost of such automated multi-camera systems. We present a networking architecture for wirelessly connected cameras and discuss research challenges and potential solutions. A particular emphasis is given on how emerging wireless technologies can be exploited towards this path and on mechanisms that could be used to relax the system requirements.**

***Keywords*-wireless communications, wireless video, IEEE 802.11, 60 GHz, video coding.**

I. Introduction

Live broadcasting of popular sport events attracts a large audience from all over the world. However, the vast majority of sport events happening every day are local, like university and school competitions. These events are of interest only to a relatively small group of viewers, and are what we call micro-events. The cost of deploying the broadcast equipment and a crew of technicians and producers is usually too high because such events are not popular enough to justify the investment.

With a camcorder and a laptop, low quality broadcasting to the Internet is technically feasible and can be offered as a commercial service [1]–[3]. More challenging is to provide a low-cost live video broadcasting of micro-events at the quality level of popular broadcasts. Recent advances in video analysis and computer vision may help provide such broadcasts: Video analysis can be used to detect relative player locations to automatically select the most appropriate camera, as well as its zooming, focus and angle factors to display to the viewers [4], [5]. The automated nature of this system minimizes human involvement in the production of the video content and hence reduces the production cost.

An automated video analysis only partially solves the problem of high investment needs because optical or Ethernet cables are still needed to transmit the high-definition (HD) video streams from each camera to the data center, and to send the steering control information from the data center to the broadcasting cameras. While professional sport stadiums are built to facilitate the deployment and cabling of multi-camera systems with flexible ducts and existing cables or dark fiber, micro-events can be at unpredictable locations, organized at pitches or fields where it might be difficult to lay cables in a reliable way. Complicated regulatory rules for construction work, listed and protected buildings adjacent to the location, or unknown site ownership may increase the investment. Temporary cabling lying on ground might be undesirable in places where vehicles could potentially destroy the installation during the event.

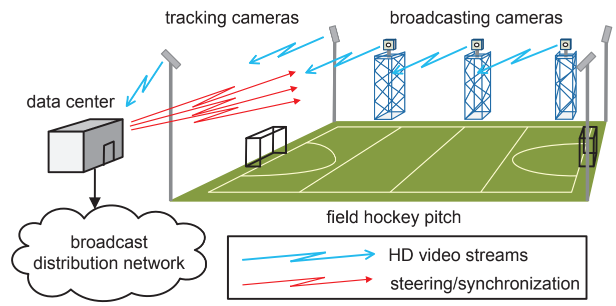

Figure 1. The microcasting system relies on tracking cameras to transmit information about the action to the central data processing center. This information is used by computer vision algorithms for automated decision making in the data center, which transmits the resulting control information to steer the broadcasting cameras. Video streams from the broadcasting cameras are sent to the data center where the final broadcast content is produced and forwarded to the content distribution network.

These factors are the main motivations for a reliable low-cost wireless networking solution. In this paper, we consider an automated multi-camera system for high-definition video broadcasting of micro-events referred to as *wireless micro-casting* and shown in Fig. 1. A set of dedicated tracking cameras is used to track the action in the field (for example the player with the ball). These cameras send wirelessly the data to a data center located next to the pitch, where decisions are made on how to steer the broadcasting cameras and when to hand over from one camera to the next. Video streams are then sent from the broadcasting cameras and distributed via a content distribution network.

978-1-4577-0351-5/11/$26.00

The deployment of the wireless microcasting is challenging: To enable the transmission of HD video streams from the cameras to the data center a wireless solution will have to support a very high throughput at low latency over distances of a few hundred meters, and be able to support a significant number of cameras/streams simultaneously. Current off-shelf technologies are unable to provide such throughputs (IEEE 802.11n offers less than 500 Mb/s) over such distances (e.g. Ultra Wideband and 60 GHz WiGig may reach few tens of meters) at reasonable cost. We discuss a number of solutions to support the required throughput and number of streams at low cost, with minimum or no loss in performance and functionality of the system.

The remainder of this paper is organized as follows: the overall system architecture of the wireless microcasting and the wireless requirements are described in Section II. Existing video compression algorithms and their impact on video quality and throughput are discussed in Section III. This motivates the discussion of wireless technologies that could be used for the network in Section IV. In Section V, we present several recent developments in video compression and computer vision that might be employed in the wireless microcasting. Section VI provides a summary of related work, and Section VII concludes the paper.

II. Wireless Microcasting - Architecture and Requirements

In the current architecture of the *microcasting* system, video cameras are connected via optical cables. Future wireless solutions need to meet several requirements to replace the cables. The microcasting architecture and the networking requirements are described in the following.

A. System Architecture

In the envisioned wireless microcasting system, illustrated in Fig. 1, there are two types of camera, one type for *broadcasting* and one for action *tracking*. Video streams from the tracking cameras, which are used for action (e.g. ball and player) tracking, are transmitted to a data processing center. An automated system in the data center analyzes the streams using computer vision algorithms and decides which broadcasting cameras, angles, and zoom factors should be used. The decisions are made in real-time and transmitted to the broadcasting cameras. Video streams from the broadcasting cameras are sent to the data processing center where the final broadcast content is produced and forwarded to the distribution network. Note that, to enable smooth handovers from one camera to the next, the frame sequences of the broadcasting cameras need to be synchronized in time using so-called generator lock (*genlock*) signal [6],

A microcasting testbed has been deployed at a field hockey pitch as shown in Fig. 2. It contains five broadcasting cameras (three at one side and two behind the goals) and eight tracking cameras. The video streams from the broadcasting cameras have a resolution of 1920×1080 pixels at 30 fps and three bytes per pixel (8-bit RGB), which gives a raw bit-rate of 1.5 Gb/s. The cameras are currently connected to the data center via fiber optic cables. The video streams from the tracking cameras have a resolution of 1920×1080 pixels at 30 fps and one byte per pixel resulting in a bit-rate of 500 Mb/s. Higher resolution for the tracking is possible and may be required by the data processing center. Currently, the bit-rate is limited by the 1 Gb/s speed of the Ethernet cables deployed on each pole that carries the tracking cameras.

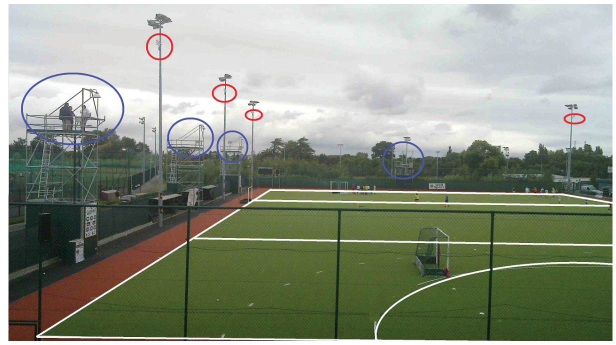

Figure 2. Reference testbed for testing at a field hockey pitch. The broadcasting cameras are mounted on the towers, while the tracking cameras are mounted on the light poles.

B. Requirements

There are several requirements that make it challenging to connect the cameras to the data center wirelessly.

High throughput. The camera network must provide high throughput to support HD video streaming. Although video sequences can often be compressed to a fraction of their original bit rate without significant loss in perceptual quality in human visions, the object detection/tracking in computer vision may be affected by compression artifacts. Therefore, none or lossless compression may be required for the video streams used by the automated decision making.

Very low latency. The information streams from the tracking cameras to the data center, and the steering/handover control data from the data center to the broadcasting cameras must be transmitted with minimum delay. This is different to most HD wireless video services, which are not interactive and therefore can tolerate a certain delay and jitter. The video streams from the broadcasting cameras to the data center may tolerate certain delays (depending on the delay introduced by the content distribution network).

Low cost. The wireless solution must also be cost-efficient compared to the wired alternatives (e.g. digging for laying optical cables). The cost of the wireless hardware is not necessarily the main deciding factor, but proprietary non-standardized technology that would incur high licensing or maintenance cost should be avoided. For example, standardized solutions enable a multi-vendor strategy.

Support of scalability. The solution must be able to scale up to a significant number of cameras/streams. Up to few tens of HD cameras can be deployed on a pitch, many of which may transmit simultaneously. ISM radio bands used by, for example, Wireless LAN (2.4 GHz ISM and 5 GHz U-NII) might not provide a sufficient number of radio channels to support so many streams. Moreover, the concurrent usage of high number of channels would increase the cost of the data center, that should be equipped with a sufficient number of wireless transceivers. Finally, other networks may use the unlicensed spectrum bandwidth, which may reduce the available bandwidth.

Moderate security. There are a few security aspects to be taken into account. A regulatory requirement in many countries is that if a video of children is captured and distributed, which can be the case in school scenarios, access to this content must be given only to educated personal and protected against open access by the public. This will require some level of wireless encryption. Further, since we look at low-cost standardized technologies, unlicensed systems such as Wireless LAN are strong candidates. Wireless LAN operates in the unlicensed Industrial, Scientific and Medical (ISM) radio spectrum and Unlicensed National Information Infrastructure (U-NII) radio spectrum which is easy to access with consumer electronics radio devices. Our final solution could therefore become susceptible to malicious activities such as intentional interference or eavesdropping.

III. Video Compression Methods to Reduce the Throughput Requirement

Ideally, video streams from the tracking and broadcasting cameras would be sent uncompressed to ensure high quality and minimum latency. However, the HD videos might need to be compressed at the cameras (e.g. using specialized hardware) to be sent wirelessly. We then investigate to what extent a reduction in the throughput requirement can be accepted, taking advantage of some popular video encoders.

Video streams transmitted by the broadcasting and tracking cameras have different quality requirements: 1) For the streams from the broadcasting cameras moderate loss in quality is acceptable since end viewers are humans with their imperfect vision. Modern video/image coding standards provide high compression efficiency: the bit-rate requirements of an uncompressed sequence can be significantly reduced at the expense of a minor loss in quality. Typically, HD videos can be losslessly compressed to a third of the original bit-rate. 2) For the streams from the tracking cameras color compression might not be desirable since even a small loss in quality may affect the performance of video analysis algorithms used for action tracking [7]. Compression also introduces latency in the camera steering/handover control signals.

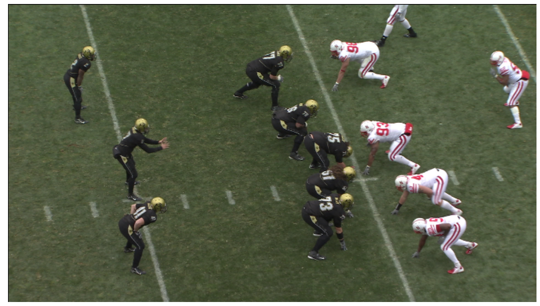

Figure 3. A screen capture from the *Touchdown Pass* test sequence.

A. Video Compression for the Broadcasting Cameras

We perform a test to estimate the compression efficiency of some popular video encoders that could be used for the wireless microcasting. We use the *Touchdown Pass* [8] test sequence (1920×1080, 30 fps, 8-bit/pixel, YUV $4:2:0$), whose screen capture is shown in Fig. 3. The bit-rate of the sequence is 750 Mb/s. The sequence is used as an example; its color-space and bit-depth may not match those produced by the broadcasting or tracking cameras used in the real life scenarios. We have encoded the sequence using two video encoders, H.264/AVC and Motion JPEG 2000.

H.264/AVC [9] is the most popular video coding standard for wireless video transmission. It provides high compression efficiency and improved error resilience mechanisms compared to previous MPEG standards. The most important shortcomings of streaming H.264/AVC videos over lossy wireless links are: i) impairments caused by transmission errors that propagate through all predicted frames, due to motion-compensation employed by the encoder, ii) bursty video traffic produced by the encoder, which poses challenges for the efficient use of radio resources. We encode the given scene sequence using constant quality (CRF) mode with a group-of-pictures (GOP) of size 30. No B-frames are used since they increase decoding complexity and introduce latency[1].

MJPEG 2000 [12] is a specification that defines the use of JPEG 2000 image compression standard for motion sequences. MJPEG 2000 does not employ motion compensation, and thus, does not propagate transmission errors through the stream. This can also be achieved with H.264/AVC using intra-coding. The main reason for considering MJPEG 2000 for wireless video streaming is the highly desirable error resilience feature. A JPEG 2000 image can be truncated at any point to obtain an image with a lower

[1]In the future, High Efficiency Video Coding (HEVC), a successor to H.264/AVC currently under development by the Joint Collaborative Team on Video Coding (JCT-VC) [10] is likely to provide a solution targeted at high- and ultra-high definition (7680×4320) video. The standard is expected to be published in 2013. Current indications are that the new standard might provide twice better video compression efficiency (i.e. around half the bit-rate for a similar quality) at the expense of significantly higher computational complexity, compared with H.264/AVC [11].

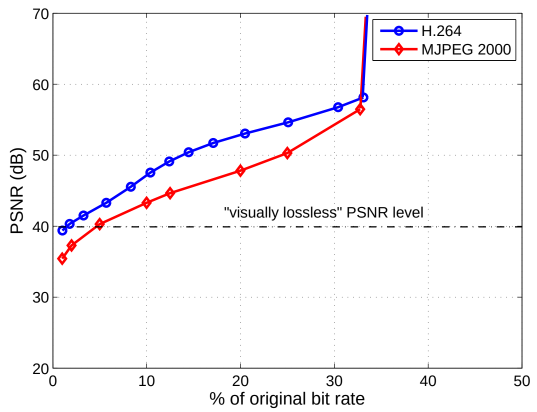

Figure 4. The impact of compression on video quality. H.264 and MJPEG 2000 are able to compress the test sequence to, respectively, 1% and 5% of its original bit-rate while keeping the PSNR at the visually lossless level.

signal-to-noise ratio. The most important shortcomings of streaming MJPEG 2000 videos are: i) lower compression efficiency compared to H.264/AVC, and ii) high computational complexity of the encoding/decoding.

We encode the video sequence with H.264/AVC and MJPEG 2000 using different compression ratios. For each resulting bit-rate, we measure the quality of the compressed sequence in terms of the average peak signal-to-noise ratio (PSNR)[2]. The PSNRs of the sequence for different compression ratios are shown in Fig. 4. With a lossless compression, PSNR is infinite since decompressed sequence is identical to the original sequence. In this case, the bit rate of the compressed sequence approximately one-third of the original bit-rate for both encoders, i.e. about 250 Mb/s in the specific sequence under test. With lossy compressions, bit rates can be further decreased at the expense of video quality. For PSNRs above 40 dB, videos are often considered to be visually lossless, i.e. a typical viewer is not able to detect the degradation in quality. According to some Mean Opinion Score (MOS) conversions, PSNRs above 37 dB are considered to be excellent (impairments are imperceptible) and PSNRs between 31 and 37 dB are good (impairments are perceptible, but not annoying) [13]. As shown in Fig. 4, H.264 and MJPEG 2000 are able to compress the test sequence close to, respectively, 1% and 5% of its original bit-rate while keeping the PSNR at the visually lossless level, corresponding to 7.5 and 37.5 Mb/s, respectively. H.264 provides significantly better compression efficiency. For example, at the PSNR equal to 50 dB, the sequence is compressed to around 15% of its original bit-rate with H.264, compared to around 25% with MJPEG 2000. However, on lossy wireless links, the superior error resilience features of MJPEG 2000 may be more beneficent than the compression efficiency of H.264. This issue requires further investigation: We plan to further evaluate the impact of video compression using video sequences from the testbed introduced in Section II-A.

B. Video Compression for the Tracking Cameras

The computer vision algorithms used in the data center for sequence analysis may require PSNR levels significantly higher than for human vision. Therefore, throughput of few hundreds of Mb/s (and thus in the level of lossless compression) might be needed on the tracking cameras to achieve those PSNR levels. To this end, we will explore different approaches: 1) We will study how compression artifacts affect the computer vision algorithms for camera steering/handover. Computer vision algorithms may be also designed to be more tolerant to such artifacts. 2) We will study how the underlying radio transmission technology can be used to support high-resolution computer vision algorithms, which is the focus of next Section.

IV. High-Speed Wireless Transmission

This Section focuses on low-cost wireless technologies that can be used for the streaming of compressed and uncompressed HD videos. We start analyzing how current technologies in the 5 GHz spectrum can be used for transmissions of visually lossless videos in broadcasting and tracking cameras. The second part of the Section will then evaluate advanced solutions in the 60 GHz band, that can be applied to uncompressed or lossless compressed HD videos for tracking cameras[3].

A. Technologies for Transmitting Visually Lossless Videos: IEEE 802.11n and Other Approaches

The IEEE 802.11 series of standards has been progressively increasing the offered data rates from 1 Mb/s in early versions up to 600 Mb/s in the existing IEEE 802.11n amendment [14]. In 802.11n, the increase is achieved with the use of up to four spatial streams over a channel width of 40 MHz. As a result, 802.11n can be used to stream visually lossless compressed HD videos.

To investigate the potential offered by current 802.11n chipset, we experimentally test the throughput using the open-source Ath9k driver for 802.11n WLM200NX Atheros chipsets. We have chosen this driver for its flexibility and reconfigurability for research scopes [15]. In the link under

[2]The PSNR measures the mean square error (MSE), pixel by pixel, of the original and compressed sequence, expressed in decibels. It is defined as $PSNR = 10 \cdot \log_{10}(MAX_i/MSE^2)$, where MAX_i is the maximum possible pixel value ($MAX_i = 255$ for 8 bits per pixel).

[3]The focus for the wireless solution lies on lowering cost with unlicensed radio technology. Alternatives such as cellular or laser links are not desirable. Laser links would be a natural choice to provide the required throughput, but require a higher investment compared to standardized radio technology that is developed for consumer markets. Further, relying mainly on standardized technology enables a multi-vendor strategy, which is important for scalability.

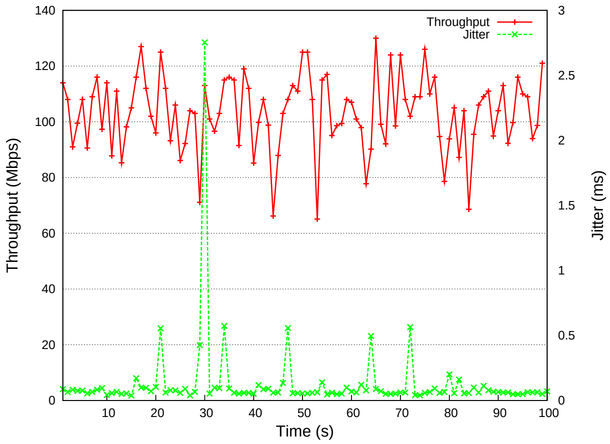

Figure 5. Throughput and delay jitter over a link with two nodes in proximity. For the test we use 802.11n WLM200NX Atheros chipsets with open-source Ath9k driver and place the node in proximity. Results show that throughputs of around 100 Mb/s are possible using commodity hardware with high flexibility for research scopes. The plot also shows that the delay jitter is mostly below 0.1 ms.

test, each station transmits and receives over two independent antennas, thus creating a Multi-Input Multi-Output (MIMO) channel, over a single-radio 40 MHz bandwidth. The nodes are placed in proximity, to guarantee a high quality link and measure the highest throughput possible. We also disable the auto-fallback algorithm to avoid any throughput reduction caused by random losses and fixed the transmission rate at a PHY rate of 300 Mb/s (the maximum one supported by the chipset). We calculate the UDP throughput and delay jitter in a test of 100 s via *iperf* [16] using the default datagram size of 1470 bytes. In order to measure the delay we also send *ping* traffic at a low rate of ten packets per second during the iperf test[4].

Results of throughput and delay jitter over time are summarized in Figure 5. The average throughput achieved in the test is 103 Mb/s, with a peak of 130 Mb/s, while the delay jitter is most of the time below 0.1 ms with an average of 0.069 ms. From the ping tests, we also find that the average round-trip-time is 11.2 ms. This throughput may be acceptable for a single HD broadcasting camera, while 802.11e prioritization may be used to guarantee a low latency (and low bandwidth) transmission for the steering data from the data center to the broadcasting camera. However, the throughput may be not sufficient for a set of HD broadcasting cameras. In fact, the deployment of multiple broadcasting cameras imply radio resources sharing and thus issues in the scalability of the solution. That is, the limited capacity of the PHY technology and the need of continuous streaming required by each camera collides with the throughput reduction due to channel contention and does not efficiently scale to a high number of cameras.

We thus expect that in 802.11n wireless microcasting only a subset of HD broadcasting cameras may be active at each time. Referring to the example in Fig. 4, for a throughput of 35 Mbps required by one camera over a visually lossless compression, we expect that at most three HD broadcasting cameras may operate concurrently. To increase the available throughput and to utilize the available throughput more efficiently, we will consider the following approach:

Link layer optimization: Many of the existing wireless technologies, including those based on the 802.11 family of standards, are oblivious to the structure of video streams. They lack proper mechanisms to provide different treatments to data packets belonging to different parts of a video stream. An uncompressed video is represented by a stream of bytes, each corresponding to the value of a given pixel. Clearly, the most significant bit (MSB) of each of these bytes has greater visual importance than the least significant bit (LSB) since an error in the LSB will result in a minor change in the pixel's value. A larger share of the available radio resources should be allocated to the important parts of the stream (e.g. by using more robust channel codes and/or modulation schemes for those parts). Such resource optimization on the link layer may help to achieve better video quality compared to general-purpose wireless technology, given the same available bit rate. The Wireless Home Digital Interface (WHDI) [17] is an example of a standard for wireless HD video transmission that employs such principles. WHDI operates in the 5 GHz band and supports data rates of up to 3 Gb/s, but only over short distances of up to 30 m.

B. Technologies for Transmitting Uncompressed and Lossless Compressed Videos: WiGig and Other Approaches

Focus of this Section is to evaluate solutions that can target uncompressed and lossless compressed videos in tracking cameras. A first option is the radio bundling of multiple 5 GHz radios. This direction has been already taken by commercial products that claim a throughput of 300 Mb/s delivered over three separated radios in outdoor links [18]. To bundle multiple radios, we imagine a scenario where a HD video is composed by multiple flows, such that each of them can be separately sent over one of the radio interfaces. An example of system that sends a single video over multiple flows was presented in [19].

Because of the limited bandwidth resources in the 5 GHz band, an interesting alternative may be provided by the upcoming standards for the 60 GHz spectrum. Activities in the IEEE 802.11ad standardization task group [14], also known as WiGig, claim to provide very high data rates required for the wireless delivery of uncompressed HD videos. The WiGig is an amendment to the 802.11 standard for operation in the 60 GHz band, where ample spectrum should be available to allow such data rates. 2010. The standard supports data rates up to 7 Gb/s at distances beyond

[4]We verified that this traffic has a negligible effect on the iperf measurement itself.

Table I
CHANNEL PARAMETERS IN OUTDOOR WIGIG

Channel bandwidth	$B = 2.16$ GHz
Center frequency	58.32, 60.48, 62.64,64.8 GHz
Wavelenght λ	5 mm at 60 GHz
EIRP	57 dBm in Europe; 40 dBm in USA
Path loss coefficient	2.5
Average distance	150 m
Noise	-74 dBm

ten meters. It defines a highly efficient directional MAC layer with the random access operations optimized for directional communication, as well as a new scheduled access mechanism[5]. A comprehensive overview of the WiGig MAC layer can be found in [21].

Our goal is to investigate whether WiGig and similar 60 GHz technologies can provide the needed throughput and latency for tracking cameras on outdoor line-of-sight links where the data center may be a hundred and more meters apart from the wireless cameras. Since WiGig is mainly designed for indoor operations, it is not clear what would be the advantage that this technology may provide in outdoor. To better understand the outdoor potential, a preliminary link budget analysis for the 60 GHz spectrum is given in the following.

60 GHz Link Budget Analysis: The link budget is an essential tool to estimate the system capacity and the trade-off between throughput and BER, based on the available bandwidth and the signal to noise ratio (SNR), which is relevant for estimating the BER at a given distance and output power or for determining the required output power or maximum distance for a target BER.

In case of WiGig, the channel bandwidth is $B = 2.16$ GHz per channel [14]. Four channels with center frequencies defined at 58.32, 60.48, 62.64, and 64.8 GHz are available. This allows to simultaneous send four streams from four different cameras. Since more cameras may be used in a stadium (like in our current setting described in Section II-A), spatial reuse is expected to be exploited via directional antennas. Systems such as WiGig must rely on high-gain directional antennas and smart beamforming techniques to increase the signal's effective radiated power and to allow the use of reflections and other indirect paths. Directional beams also allow better frequency reuse, which is desirable when many cameras stream simultaneously to the data center.

While the directionality of the wireless cameras is mostly determined by the spatial reuse required by the system, the antenna gain at the data center receiver depends on the radio propagation characteristics of millimeter waves [22]. In fact, there is no strict requirement of directionality of the steering control data from the data center, and signals may be wirelessly broadcasted to the set of broadcasting cameras, that may then decide what piece of data packet is directed to them. We are thus interested in understanding what type of antenna gains are needed to compensate for the losses. For a target of $SNR = 10$ dB at the receiver, the Shannon capacity is equal to:

$$C = B\, log_2(1 + SNR) = 7.4\ Gb/s. \tag{1}$$

We can then express SNR as:

$$SNR = EIRP + Gr - PL(d) - N - W, \tag{2}$$

where $EIRP$ is the *Equivalent Isotropically Radiated Power*, Gr is the receiver gain, $PL(d)$ is the path loss at distance d, N is the noise power at the receiver, and W is the link margin. According to the 802.11ad standard draft, the $EIRP$ is 57 dBm in Europe and 40 dBm in USA.

While no obstruction is expected between the transmitter and receiver (and thus no shadowing), a good link margin may deal with different effects. For example, radio waves in this band are usually strongly attenuated by the atmosphere and particles contained in it. Furthermore, in frequencies around 60 GHz, the radio waves are strongly attenuated by molecular oxygen in the atmosphere. Thus, an attenuation due to oxygen molecules (causing an extra attenuation of up to 15 dB/km) and to the rain (that can also be of some importance in the mm-wave band) may be added to the path loss. However, for the specific case of wireless microcasting, where we expect ranges less than 200 m, oxygen and rain attenuations can be mostly neglected [22]. Also some interference may occur between independent transmit cameras, which may reduce the SNR at the receiver. To take into account these effects, we consider a link margin of $W = 10$ dB.

Regarding the noise power N, it can be calculated as:

$$N = 10 \log_{10}(KTB) + NF =$$
$$-174\text{dBm/Hz} + 10\ \log_{10} B + NF = -74\ \text{dBm},$$

if the main source of noise is the thermal power, and where NF is the noise figure, that we suppose equal to 6 dB and KT is the noise power spectral density, i.e., the product of Boltzmann constant with temperature.

Since we expect an unobstructed LOS between the antenna and the receiving unit, under the hypothesis that transmit and receive antennas have the same physical orientation to match the polarization, the signal follows the Friis equation $PL_0(d) = \left((4\pi d)/\lambda\right)^2$ until a breakpoint distance d_0, where $\lambda = 5\ mm$ at 60 GHz indicates the wavelength. The path loss is a function of λ, and as a result the higher frequency of WiGig causes a higher attenuation respect to

[5]WirelessHD [20] is another recent technology that operates in the 60 GHz range. It is an industry-defined specification for wireless HD video/audio transmission for consumer electronics. WirelessHD aims to support significantly higher data rates compared to the WiGig (up to 28 Gb/s), but over shorter distances (up to 10 m). Essentially, the WirelessHD is designed as a wireless equivalent of HDMI and, therefore, it is not a direct competitor to the WiGig.

2.4 and 5 GHz Wi-Fi. Assuming a breakpoint distance of $d_0 = 10$ m [23], $PL_0(d_0 = 10\,m) = 20\log_{10}((4\pi 10)/\lambda) = 88$ dB. Above 10 meters, the outdoor channel is close to the free space loss channel, with path loss coefficient reported to be up to $n = 2.5$ [22]. Then, for $d \geq d_0 = 10\,m$, the path loss can be expressed as:

$$PL(d) = PL_0(d_0) + 10n\,\log_{10}(d) - 10n\,\log_{10}(d_0).$$

If we consider a target distance of $d = 150$ m to guarantee enough coverage range in the stadium (that is the maximum distance between a tracking camera and the data center) at the target SNR, we need to add an attenuation of $25\log_{10} 15 = 30$ dB. Thus:

$$PL(150\,m) = 88 + 30 = 118 \text{ dB}$$

Thus, from equation (2), a receiver antenna gain of $Gr \geq 10 - 40 + 118 - 74 + 10 = 24$ dBi is needed in USA, while in Europe: $Gr \geq 10 - 57 + 118 - 74 + 10 = 7$ dBi. Concluding, WiGig can potentially work in outdoor links in wireless microcasting with directional antennas.

We aim to further investigate the accuracy of the path loss model and verify with experimental hardware if a free space path loss model ($n = 2$) can be also applied in our deployment, that would translate in a path loss $PL(d = 150\,m) = 111$ dB. We are also interested in measuring the cross-interference at the data center among multiple WiGig signals with different transmit antenna gains.

V. Additional Approaches for Low-Cost Video Networking

There are ways to relax the overall radio resource requirements of our system, some of which we are considering for future exploration. In this Section we discuss such ways and advanced solutions. For example, there are known methods to mitigate the need for additional tracking cameras by using all the broadcasting cameras for the tracking instead. This would limit the amount of data transmitted over the wireless channels. Smart camera concepts may help by transmitting data and tracking information separately with different levels of quality-of-service and thus staying in the required time limits for the decision making and camera control (Section V-A). The overall decision making could exploit the focus of the audience and visitors, as discussed in Section V-B. New concepts based on distributed video may help exploiting redundancy and correlation between the different video streams, which is highlighted in Section V-C. Graceful degradation of video quality in the presence of throughput variations can be achieved with the scalable video coding, as discussed in Section V-D.

A. Smart Cameras

Smart cameras are cameras capable of on-board video processing for scene analysis and metadata extraction [24]. Smart tracking cameras could be used to extract relevant information from uncompressed video streams by means of background subtraction, blob forming, and object tracking algorithms. Only the important portions of the video are transmitted to the data center to significantly reduce the throughput requirements for the tracking cameras. Furthermore, video processing and action tracking might be performed on smart broadcasting cameras, thus eliminating the need for dedicated tracking cameras. A prototype of a camera that computes and transmits the tracking information only to enable real-time tracking of objects and persons is presented in [25].

B. Audience Tracking

In the microcasting system, camera steering/handover algorithms analyze the position and movement of players and a ball (in case of ball games) based of the streams from the tracking cameras in order to derive targets for the broadcasting cameras. However, the target locations could also be derived, not from the complex action on the field, but from the behavior of the audience around the field. The tracking cameras could observe the audience to estimate its focus of attention using head pose estimation, gaze direction estimation, and similar techniques. For example, in [26], the authors propose an automatic pan control system that tracks face direction of the audience. Potential benefits of this approach are: i) the number of tracking cameras, thus video streams, can be reduced since the behavior of the audience is often homogeneous and ii) the algorithm may be more tolerant compression artifacts, thus allowing higher compression ratios, assuming that the tracking cameras are placed close to the audience.

C. Distributed Video Coding

Adjacent cameras have partially overlapping views of the pitch. Therefore, some of the video streams are highly correlated, but joint encoding is not practical since it would require wireless communication between the cameras. Fortunately, distributed video coding (DVC) addresses scenarios where multiple correlated video sequences are separately encoded, but jointly decoded (at the data center), thus not requiring any communication between the cameras. The DVC is based on two major results from information theory, the Slepian-Wolf [27] and Wyner-Ziv [28] theorems, which suggests that the minimum rate to separately encode two correlated sources (X and Y) with an arbitrary small probability of error is the same as the minimum rate for joint encoding, when joint decoding is performed and the difference $X - Y$ is Gaussian distributed.

Based on the two theorems, several algorithms for distributed video coding (DVC) have been proposed. For example, the algorithm proposed in [29] has been adopted for the DVC video codecs developed in the context of VISNET [30], and DISCOVER [31] projects. However, practical DVC algorithms are still in an infancy stage.

We will explore the feasibility of the DVC and propose needed improvements to the existing algorithms for microcasting scenarios. The potential benefits of the DVC for the microcasting are i) reduction of the transmission rates by exploiting the correlation between camera views, ii) reduced encoding complexity in the resource-limited cameras at the expense of more complex joint decoding in the data center, and iii) improved resilience to channel errors since DVC facilitates joint source-channel coding.

D. Scalable Video Coding

Scalable Video Coding (SVC) is a coding method that enables multiple versions of a video (e.g. versions with different qualities, frame rates, and resolutions) to be stored in one stream by encoding the difference between the versions. A reference implementation of the SVC is an extension of the H.264/AVC video coding standard [32]. An SVC video stream consists of a single base layer, which must be available at the receiver in order to decode the lowest quality version, and a number of enhancement layers. The SVC provides graceful degradation in video quality when link capacity varies over time since it allows the receiver to extract a video from a subset of received layers. Appropriate link layer mechanisms are needed to provide different levels of robustness to link errors to different layers. Hence, mechanisms such as the unequal error protection (UEP) and hierarchical modulation are often considered in the context of scalable video transmission. Note that adding and dropping of the enhancement layers based on the currently available throughput introduces annoying fluctuations in video quality, which are not acceptable for high-quality video broadcasting, but might be for the microcasting. Nevertheless, the underlaying wireless solution should be designed so that the scalability mechanisms of the SVC are used only as a fail-safe in case of occasional disturbances in link quality, rather than as mechanisms on which the system relies continuously.

VI. Related Work

In [4], the authors explore the design of an automated computer-driven sports broadcast director that provides personalized automated broadcasts, depending on the viewers preferences as well as the specific actions unfolding in the game. The system uses video analysis to detect the relative player locations to automatically select the most interesting camera angle to display to that viewer in real-time. Similarly, [5] describes an approach for an automatic sports director for hockey. Stationary cameras are used to track the hockey players over the course of a game and then generate an automatic broadcast from the resulting data gathered. The broadcast content however is not generated in real-time. The automated sports broadcast approach for the soccer game described in [33] depends on higher-level, sport-specific semantic information, such as a shot or a foul recognized from audio and commentary, in order to determine which camera angle to show. [34] describes a system that uses received signal strength data from multiple strategically placed sensor nodes to localize the game assets (e.g. ball, players) and automate the control of broadcasting cameras. FoxTrax [35] adopted a similar approach for real-time tracking of an ice-hockey puck in a game.

An overview of wireless technology capable of streaming compressed and uncompressed high-definition videos is provided in [36]. The feasibility of the HD video transmission over short distances (up to few meters) using the ultra-wideband (UWB) technology is analyzed in [37], [38]. A design of a 60 GHz transceiver chipset capable of streaming uncompressed $1080\mathrm{p}/60$ videos at distances of up to ten meters is described in [39]. In [40], the authors present a 60 GHz system that supports uncompressed HD videos with data rates of up to 3 Gb/s. The system includes error protection and concealment schemes that exploit unequal error resilience properties of uncompressed video. A system based on IEEE 802.11ac that operates in 5 GHz band with 80 MHz bandwidth and provides bit-rates above 1.5 Gb/s is presented in [41]. In this system, video is compressed by MJPEG 2000 and uses its advanced error resilience tools. In [42], the authors propose an error correction scheme for wireless video transmission that uses the large amount of spatial redundancy already present in uncompressed HD video to provide an extra layer of protection in addition to that provided by channel coding.

VII. Conclusion

In this paper, we discussed wireless microcasting, an automated wireless system for broadcasting of live sport events at a lower cost compared to current streaming solutions. We described the research challenges posed by microcasting and the requirements that must be met to build the system. As an outcome of this discussion, it emerges that only a close-loop interaction among wireless technologies, video coding and computer vision can offer the potential to build such low cost solution. We aim to further explore this interaction with real data from the hockey pitch testbed.

Acknowledgments

We thank Andreas F. Molisch, Professor of Electrical Engineering at the University of Southern California, Los Angeles, USA, and Thomas Gross, Professor of Computer Science Department at ETH, Zurich, Switzerland, for their valuable comments provided to this contribution.

References

[1] "Qrodo." http://www.qrodo.com.

[2] "Livestream." http://www.livestream.com.

[3] "YouTube Live." http://www.youtube.com/live.

[4] M. Chuang and P. Narasimhan, "Automated viewer-centric personalized sports broadcast," *Procedia Engineering*, vol. 2, no. 2, pp. 3397 – 3403, 2010.

[5] L. Wu, "Multi-view hockey tracking with trajectory smoothing and camera selection," *Master Thesis, The University Of British Columbia*, 2008.

[6] "An Overview of Genlock." http://www.mivs.com/technical/appnotes/an005.html.

[7] G. J. Burghouts and J.-M. Geusebroek, "Performance evaluation of local colour invariants," *Comput. Vis. Image Underst.*, vol. 113, pp. 48–62, January 2009.

[8] "Touchdown Pass video sequence." Available at: http://media.xiph.org/video/derf/.

[9] "ITU Rec. H.264/ISO IEC 14996-10: Advanced Video Coding for Generic Audiovisual Services." http://www.itu.int/rec/T-REC-H.264/.

[10] "Joint Collaborative Team on Video Coding (JCT-VC)." http://www.itu.int/en/ITU-T/studygroups/com16/video/.

[11] G. J. Sullivan and J.-R. Ohm, "Recent developments in standardization of high efficiency video coding (HEVC)," in *Applications of Digital Image Processing XXXIII*, vol. 7798, aug 2010.

[12] "ITU-T T.802/ ISO IEC 15444-3: Information technology JPEG 2000 image coding system: Motion JPEG 2000." http://www.itu.int/rec/T-REC-T.802/.

[13] J. Klaue, B. Rathke, and A. Wolisz, "Evalvid - a framework for video transmission and quality evaluation," in *Proc. 13th Int. Conf. Modelling Techniques and Tools for Computer Performance Evaluation*, (Chicago, USA), 2003.

[14] "IEEE 802.11 working group for wireless LANs." http://www.ieee802.org/11/.

[15] "Linux Wireless." http://wireless.kernel.org/.

[16] "Iperf." http://sourceforge.net/projects/iperf/.

[17] "WHDI Wireless High Definition." http://www.whdi.org/.

[18] "Meraki MR58." http://meraki.com/products_services/access_points/MR58/.

[19] X. Yang, M. Gjoka, P. Chhabra, A. Markopoulou, and P. Rodriguez, "Kangaroo: video seeking in p2p systems," in *Proc. 8th Int. Conf. on Peer-to-Peer Systems*, IPTPS'09, (Boston, USA), 2009.

[20] "WirelessHD." http://www.wirelesshd.org/.

[21] C. Cordeiro, D. Akhmetov, and M. Park, "Ieee 802.11ad: introduction and performance evaluation of the first multi-gbps wifi technology," in *Proc. ACM Int. Workshop on mmWave Communications: From Circuits to Networks*, mmCom '10, (New York, USA), 2010.

[22] P. Smulders and L. Correia, "Characterisation of propagation in 60 ghz radio channels," in *Electronics and Communication Engineering Journal*, pp. 73–80, 1997.

[23] E. Grass, F. Herzel, M. Piz, Y. Sun, and R. Kraemer, "Implementation aspects of gbit/s communication system for 60 ghz band," *14th Wireless World Research Forum (WWRF 14)*, July 2005.

[24] W. Wolf, B. Ozer, and T. Lv, "Smart cameras as embedded systems," *Computer*, vol. 35, pp. 48–53, Sep. 2002.

[25] S. Fleck, S. Lanwer, and W. Straer, "A smart camera approach to real-time tracking," in *Proc. 13th European Signal Processing Conference*, EUSIPCO, (Istanbul, Turkey), 2005.

[26] S. Daigo and S. Ozawa, "Automatic pan control system for broadcasting ball games based on audience's face direction," in *Proc. ACM Int. Conf. Multimedia*, (New York, NY, USA), 2004.

[27] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inform. Theory*, vol. 19, no. 4, pp. 471–480, 2003.

[28] A. Wyner, "Recent results in Shannon theory," *IEEE Trans. Inform. Theory*, vol. IT-20, no. 2, pp. 2–10, 1974.

[29] A. Aaron and B. Girod, "Compression with side information using turbo codes," in *Proc. IEEE Data Compression Conf.*, DCC '02, (Snowbird, USA), pp. 252–261, 2002.

[30] "VISNET I/II: Networked Audiovisual Media Technologies." http://www.visnet-noe.org/.

[31] "DISCOVER: Distributed Coding for Video Services." http://www.discoverdvc.org/.

[32] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the Scalable Video Coding Extension of the H.264/AVC Standard," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 17, pp. 1103–1120, Sep. 2007.

[33] J. Wang, C. Xu, E. Chng, H. Lu, and Q. Tian, "Automatic composition of broadcast sports video," *Springer Multimedia Systems*, vol. 14, no. 4, pp. 179–193, 2008.

[34] H. Gandhi, M. Collins, M. Chuang, and P. Narasimhan, "Real-time tracking of game assets in american football for automated camera selection and motion capture," *Procedia Engineering*, vol. 2, no. 2, pp. 2667 – 2673, 2010.

[35] R. Cavallaro, "The foxtrax hockey puck tracking system," *IEEE Computer Graphics Applications*, vol. 17, pp. 6–12, March 1997.

[36] G. Lawton, "Wireless hd video heats up," *IEEE Computer*, vol. 41, no. 12, pp. 18–20, 2008.

[37] R. Ruby, Y. Liu, and J. Pan, "Evaluating video streaming over uwb wireless networks," in *Proc. 4th ACM Workshop on Wireless multimedia networking and performance modeling*, WMuNeP '08, (Vancouver, Canada), 2008.

[38] D. Porcino, B. van der Wal, and Y. Zhao, "Hdtv over uwb: wireless video streaming trials and quality of service analysis," *JPEG2000 Technical article, Analog Devices, Inc.*, 2006.

[39] J. M. Gilbert, C. H. Doan, S. Emami, and C. B. Shung, "A 4-gbps uncompressed wireless hd a/v transceiver chipset," *IEEE Micro*, vol. 28, no. 2, pp. 56–64, 2008.

[40] H. Singh, J. Oh, C. Kweon, X. Qin, H.-R. Shao, and C. Ngo, "A 60 ghz wireless network for enabling uncompressed video communication," *IEEE Communications Magazine*, vol. 46, no. 12, pp. 71–78, 2008.

[41] M. Kurosaki, Y. Hirata, M. Matsuo, W. A. Syafei, B. Sai, Y. Kuroki, A. Miyazaki, and H. Ochi, "A study on wireless transmission system for full-spec hdtv," in *Proc. 11th Int. Conf. Advanced Communication Technology*, ICACT'09, (Phoenix Park, USA), 2009.

[42] M. Manohara, R. Mudumbai, J. Gibson, and U. Madhow, "Error correction scheme for uncompressed hd video over wireless," in *Proc. IEEE Int. Conf. Multimedia and Expo*, ICME'09, (New York, USA), 2009.