A Remote Markerless Human Gait Tracking for E-Healthcare Based on Content-Aware Wireless Multimedia Communications

Haiyan Luo  
*University of Nebraska-Lincoln*

Song Ci  
*University of Nebraska-Lincoln*

Dalei Wu  
*University of Nebraska-Lincoln*

Nicholas Stergiou  
*University of Nebraska at Omaha, nstergiou@unomaha.edu*

Ka-Chun Siu  
*University of Nebraska Medical Center*

Follow this and additional works at: [https://digitalcommons.unomaha.edu/biomechanicsarticles](https://digitalcommons.unomaha.edu/biomechanicsarticles)

Part of the Biomechanics Commons

Recommended Citation

Luo, Haiyan; Ci, Song; Wu, Dalei; Stergiou, Nicholas; and Siu, Ka-Chun, "A Remote Markerless Human Gait Tracking for E-Healthcare Based on Content-Aware Wireless Multimedia Communications" (2010). *Journal Articles*. 110.  
[https://digitalcommons.unomaha.edu/biomechanicsarticles/110](https://digitalcommons.unomaha.edu/biomechanicsarticles/110)
A Remote Markerless Human Gait Tracking for E-Healthcare Based on Content-Aware Wireless Multimedia Communications

Haiyan Luo, Song Ci, and Dalei Wu
University of Nebraska-Lincoln

Nicholas Stergiou
Nebraska Biomechanics Core Facility

Ka-Chun Siu
University of Nebraska Medical Center

Abstract

Remote human motion tracking and gait analysis over wireless networks can be used for various e-healthcare systems for fast medical prognosis and diagnosis. However, most existing gait tracking systems rely on expensive equipment and take lengthy processes to collect gait data in a dedicated biomechanical environment, limiting their accessibility to small clinics located in remote areas. In this work we propose a new accurate and cost-effective e-healthcare system for fast human gait tracking over wireless networks, where gait data can be collected by using advanced video content analysis techniques with low-cost cameras in a general clinic environment. Furthermore, based on video content analysis, the extracted human motion region is coded, transmitted, and protected in video encoding with a higher priority against the insignificant background area to cope with limited communication bandwidth. In this way the encoder behavior and the modulation and coding scheme are jointly optimized in a holistic way to achieve the best user-perceived video quality over wireless networks. Experimental results using H.264/AVC demonstrate the validity and efficacy of the proposed system.

Introduction

It is reported that certain diseases can directly affect human gait, stride, pace, and overall walking pattern. Human gait tracking, the process of identifying an individual by walking manner, can then be widely used in surveillance, biomedical assistance, physical training and therapy. For instance, it can readily be used to provide prognostic and diagnostic measures of pathological locomotion biorhythm such as Parkinson’s disease, diabetic peripheral neuropathy, and Huntington’s disease. It can also be utilized for the clinical assessment of stroke rehabilitation, prosthetic alignment, and the success of orthopedic interventions such as anterior cruciate ligament reconstruction. Therefore, human motion tracking and gait analysis create an opportunity for early intervention and possibly the prevention of injury as a result of declining condition from disease or the natural aging process [1, 2]. However, most existing tracking systems usually rely on expensive equipment and lengthy processes to collect gait data in a dedicated biomechanical environment, limiting their accessibility to small clinics located in remote areas.

On the other hand, the increasing prevalence of inexpensive hardware such as video phones. And complementary metal oxide semiconductor (CMOS) cameras has fostered the development of wireless
multimedia technologies, such as wireless multimedia sensor networks (WMSNs). WMSNs not only enhance existing sensor network applications such as tracking, surveillance, home automation, and environmental monitoring, but also facilitate new medical applications such as telemedicine and advanced healthcare delivery. Indeed, telemedicine is a fast-growing application of clinical medicine, where medical information is transferred via telephone, the Internet, or wireless networks for the purpose of consulting, and sometimes remote medical procedures or examinations. Telemedicine can be integrated with third-generation (3G) broadband multimedia networks to provide ubiquitous e-healthcare services.

Furthermore, remote monitoring is a new technology emerging to improve disease treatment and lower medical costs. The essence of remote monitoring is to enable assessment of an individual’s medical status in real time regardless of his or her location, and to allow a doctor or a computer to view the information anywhere to aid diagnosis, observe how a treatment is working, or determine if a condition has become acute. By reducing the number of visits to clinics or care facilities, medical costs can be significantly reduced, and convenience and care quality are highly improved.

Therefore, an e-healthcare platform based on wireless multimedia technologies for remote markerless human motion tracking and gait analysis will significantly improve current clinical medicine practices, especially for small clinics located in remote areas. Patients in remote areas can also get online screening or a preliminary diagnosis before they are physically transported to a fully equipped medical center to save time and resources.

In this article we present a new markerless human gait tracking based on content-aware wireless streaming for remote tracking, as shown in Fig. 1. The proposed system first extracts the human motion region accurately by jointly utilizing properties of interframe relations, as well as spectral and spatial inter-pixel-dependent contexts of the video frames. Thus, it does not need the expensive equipment in the traditional pre-designed intrusive marker-based environment (Fig. 2). Then, based on content-aware analysis results, collected gait data are transmitted to the medical center for speedy or real-time medical prognosis and diagnosis through wireless networks. A proposed quality-driven distortion-delay framework is adopted to guarantee user-perceived video quality at the receiver end. Specifically, video coding parameters at the application layer, and a modulation and coding scheme at the physical layer are jointly optimized through a minimum distortion problem to satisfy a given playback delay deadline. In other words, the optimal combination of encoder parameters and adaptive modulation and coding (AMC) scheme are chosen to achieve the best fidelity of the collected gait data. Also, the extracted human gait region can be coded with finer parameters, and be transmitted with better protection than the insignificant background area under the given quality of service (QoS) requirement. Experimental results using H.264/AVC have shown the validity and effectiveness of the proposed system.

The remainder of this article is organized as follows. We present the procedures for markerless human motion tracking. We also discuss the proposed optimized delivery of wireless streaming based on the content-aware analysis results of the previous section. We then describe the system environment as well as the experimental results. Finally, we conclude the article.
Markerless Human Gait Tracking

Given a gait video sequence, the human gait region contains valuable information for gait analysis and medical examinations. In contrast, the background area usually provides much less useful knowledge. However, according to our study, for a video frame of the collected human gait data, the background area usually comprises more than 50 percent of the whole video frame area. Transmitting the background area and the human gait region without differentiation is not only unnecessary but also wasteful of precious wireless network resources. Therefore, the identification of human gait region through the proposed content-aware analysis techniques is a meaningful task. The result provides a foundation for wireless transmission of remote human gait tracking by prioritizing the human gait region against the unimportant background area under the bandwidth-limited environment. Furthermore, the proposed marker-less human motion tracking system makes gait tracking much less dependent on dedicated gait tracking facilities.

Figure 1. The application scenario of the proposed quality-driven content-aware wireless streaming for remote markerless human gait tracking.
Figure 3 shows the proposed procedures for markerless human gait tracking and the corresponding results, where background subtraction and contextual classification play major roles. Through these two steps, the temporal inter-frame correlation of the video clip, the spectral and spatial inter-pixel-dependent contexts of the video frames are cooperatively utilized, making it possible to accurately track the human gait region in a markerless environment.

**Background Subtraction** - The rationale of background subtraction is to detect the moving objects from the difference between the current frame and a reference frame, often called the "background image." Therefore, the background image is ideally a representation of the scene with no moving objects and is kept regularly updated so as to adapt to the varying luminance conditions and geometry settings [3]. Given a video sequence, the objective of background subtraction is to detect all foreground objects (i.e., the human gait region in this article). The naive description of the approach is to depict the human gait region as the difference between the current frame $F_r$ and the background image $B_g$: 
\[|Fr_i - B_{fg}| > T_h, \quad (1)\]

where \(T_h\) denotes a threshold.

However, the background image is not totally static. Therefore, before this approach can actually work, certain factors need to be adapted to, which includes illumination changes, motion changes, as well as sometimes changes in background geometry. Over time, different background objects are likely to appear at the same pixel location. Sometimes the changes in the background object are not permanent and appear at a rate faster than that of the background update. To model this scenario, a multivalued background model can be adopted to cope with multiple background objects. Therefore, algorithms such as the proposed Gaussian Mixture Model (GMM) can define an image model more properly as it provides a description of both foreground and background values [4]. The result of this step using GMM is illustrated in Fig. 3b.

**Contextual Classification**: The objective of classification is to classify a video frame by the object categories it contains. Supervised classification is a type of automatic multispectral image interpretation in which the user supervises feature classification by setting up prototypes (collections of sample points) for each feature class to be mapped. A supervised contextual classification that utilizes both spectral and spatial contextual information can better discriminate between pixels with similar spectral attributes but located in different regions. First, in many images, especially those remotely sensed images, object sizes are much greater than the pixel element size. Therefore, the neighboring pixels are more likely to belong to the same class, forming a homogeneous region. Furthermore, some classes have a higher possibility of being placed adjacent to others, so the information available from the relative assignments of the classes of neighboring pixels is also very important. By using both spectral and spatial contextual information, the speckle error can effectively be reduced, and the classification performance can be improved significantly. Nonetheless, this type of classification also suffers from the problem of small training sample size, where the class conditional probability has to be estimated in the analysis of hyperspectral data.

Therefore algorithms such as adaptive Bayesian contextual classification that utilizes both spectral and spatial inter-pixel-dependent contexts to estimate the statistics and classification can be adopted for accurate classification [5]. This model is essentially the combination of a Bayesian contextual classification and an adaptive classification procedure. In this classification model only inter-pixel class dependency context is considered, while the joint prior probabilities of the classes of each pixel and its spatial neighbors are modeled using Markov random fields (MRF) [6]. The result of this step is shown in Fig. 3c.

**Region Growing**: At this stage, shown in Fig. 3d, density check is adopted to combine the results of the previous two steps to form the continuous human gait region. As long as a homogeneous region achieved from stage 2 contains more than a threshold percentage of human motion pixels obtained from stage 1, this region is regarded as the human gait region. Otherwise, it falls into the background area.

**Morphological Operations and Geometric Corrections**: Results from the previous stage contain undesired noises and holes. As shown in Fig. 3e, morphological operations use dilation and erosion to
Wireless Streaming of Remote Human Gait Tracking

The Proposed System Model

Based on the proposed markerless human motion tracking results, we also develop a unified quality-driven optimization system of wireless streaming for delay-bounded human gait transmission. Figure 4 illustrates the proposed system model, which consists of an optimization controller, the marker-less human gait tracking module, a video encoding module, as well as the modulation and coding module. To increase the overall video quality, we first adopt the proposed methods mentioned earlier to identify the human gait region in a marker-less environment. Due to the different contribution to gait analysis from the human motion region and the background area, the human gait region can be coded at finer quality at the video encoder of the application layer, and the packets of the human gait region can use smaller constellation sizes and lower channel coding rates to guarantee the required packet error rate at the physical layer. By redistributing the limited resources needed for encoding and transmission according to the video content, the overall quality of real-time human gait video delivery over wireless networks will be significantly improved.

Furthermore, the controller is the core of the proposed system, which can jointly optimize the key system parameters of the video codec at the application layer, and the modulation and coding schemes at the physical layer. Therefore, through these parameters, the controller can control the behaviors of the video encoder and the modulation and coding module. More important, adjusting with coordination the system parameters of the video encoder residing in the application layer and the modulation and coding scheme residing in the physical layer can greatly enhance overall network performance. As shown in Fig. 5, the system performance in terms of video distortion is jointly decided by the encoder behavior (i.e., quantization step size, or QP) and the packet loss rate. Additionally, packet loss rate is determined by bit error rate (BER), which is then collectively affected by the channel quality and AMC scheme. Therefore, all related system parameters can be holistically optimized toward achieving the best possible video quality under a given delay constraint. For example, when a wireless channel is experiencing bad quality, the time-varying channel information can be used to dynamically adapt the AMC scheme to minimize the packet loss rate, enhancing the received video quality over wireless networks. Thus, the proposed cross-layer joint optimization is able to choose the optimal set of parameter values to achieve the best received video performance, providing a natural solution to improve overall system performance for wireless streaming of remote human gait tracking.

The Joint Optimization of Content-Aware Wireless Streaming

At the video encoder, for hybrid motion-compensated video coding and transmission over lossy channels, each video frame is generally represented in block-shaped units of the associated luminance and chrominance samples (16 x 16 pixel region) called macroblocks (MBs). In the H.264 codec MBs can
be either intra-coded or inter-coded from samples of previous frames [8]. Intra-coding is performed in the spatial domain, by referring to neighboring samples of previously coded blocks to the left and/or above the block to be predicted. Inter-coding is performed with temporal prediction from samples of previous frames. Many coding options exist for a single MB, and each of them provides different rate-distortion characteristics. In this work only pre-defined MB encoding modes are considered, since we want to apply error-resilient source coding by selecting the encoding mode of each particular MB. This is crucial to allow the encoder to trade off bit rate with error resilience at the MB level. For real-time source coding, the estimated distortion caused by quantization, packet loss, and error concealment at the encoder can be calculated by using the Recursive Optimal Per-Pixel Estimate (ROPE) method, which provides an accurate video-quality-based optimization metric to the cross-layer optimization controller [9].

At the physical layer, the BER $p_m^e$ is decided by the dynamically chosen mode of AMC, which has been advocated to enhance the throughput of future wireless communication systems at the physical layer [10]. With AMC, the combination of different constellations of modulation and different rates of error control codes are chosen based on the time-varying channel quality. For example, in good channel conditions, an AMC scheme with larger constellation sizes and high channel coding rate can guarantee the required packet error rate, which means that AMC can effectively decrease the transmission delay while satisfying the packet loss rate constraining. Each AMC mode consists of a pair of modulation scheme $\alpha$ and FEC code $c$ as in Third Generation Partnership Project (3GPP), HIPERLAN/2, IEEE 802.11a, and IEEE 802.16 standards. Furthermore, we adopt the following approximated BER expression:

$$p_m^e(\gamma) = \frac{a_m}{e^{b_m} \gamma},$$

(2)

where $m$ is the AMC mode index and $\gamma$ is the received SNR. Coefficients $a_m$ and $b_m$ are obtained by fitting Eq. 2 to the exact BER as shown in Fig. 5.

Therefore, the expected mean squared error (MSE) between the received pixels and original pixels of the video frames can be adopted as the

![Figure 1: The different system parameters and their interactions that are considered in the joint optimization of wireless streaming for remote human gait tracking.](image)
Thus, the expected distortion \( E_d(\rho) \) accurately calculated by ROPE under instantaneous network conditions, which is represented by packet loss rate \( \rho \), becomes the objective function in the proposed optimization framework. Packet loss rate \( \rho \) can be further calculated from BER \( p_m \) as long as the packet size is known. Meanwhile, the transmission delay \( t_{trans} \) which is constrained by the given frame delay bound \( T_f \), can be represented by bandwidth \( B_e \) and data bit rate \( R_m \). Finally, the problem can be formulated as a minimum distortion problem constrained by a given frame delay bound.

\[
\begin{align*}
\text{minimize} & \quad E_d(\rho) \\
\text{subject to} & \quad t_{trans}(B_e,R_m) < T_f
\end{align*}
\]

By eliminating from the potential solution set the parameters that make the transmission delay exceed the delay constraint, the constrained problem can be relaxed to an unconstrained optimization problem. Furthermore, most decoder concealment strategies introduce dependencies among slices. For example, if the concealment algorithm uses the motion vector of the previous MB to conceal the lost MB, it would cause the calculation of the expected distortion of the current slice to depend on its previous slices. Therefore, we can assume that the current slice depends on its previous \( z \) slices (\( z \geq 0 \)). Then, given the current decision vectors, the selection of the next decision vector is independent of the selection of the previous decision vectors, which makes the future step of the optimization process independent of its past steps, forming the foundation of dynamic programming. Thus, the problem can be converted into and solved as a well-known problem of finding the shortest path in a weighted directed acyclic graph (DAG) \[11\]. In this way the optimization problem is efficiently solved \[12\].

**System Experiments**

In the experiments video coding is performed by using the H.264/AVC JM 12.2 codec, where the gait video is recorded through an ordinary video camera in an indoor environment, as shown in Fig. 3. The frames of the recorded QCIF sequence are coded at the frame rate \( R_{frame} \) of 30 frames/s, where each \( I \)-frame is followed by 9 \( P \)-frames. We set one packet to be one slice (one row of MBs). When a packet is lost during transmission, we use the temporal-replacement error concealment strategy. The motion vector of a missing MB is estimated as the median of motion vectors of the nearest three MBs in the preceding row. If that row is also lost, the estimated motion vector is set to zero. The pixels in the previous frame, pointed to by the estimated motion vector, are used to replace the missing pixels in the current frame. Furthermore, we adopt the Rayleigh channel model to describe signal-to-noise ratio (SNR) \( \gamma \) statistically. For channel adaption, we assume the channel is frequency flat, remaining time invariant during a packet, but varying from packet to packet. Therefore, AMC is adjusted on a packet-by-packet basis. Besides, we also adopt perfect channel state information at the receiver, which is fed back to the transmitter without error or latency. For the joint optimization of QP and AMC, we allow QP to range from 1 to 50 and AMC to be chosen from the six available schemes in Fig. 5. We use NS-2 and MATLAB for system experiments. A serial multi-hop network topology is adopted using NS-2. The bandwidth and the propagation delay on each link are fixedly set to 106 symbols/s and 10 \( \mu \)s, respectively.

The expected video quality at the receiver is measured by the average of the peak SNR (PSNR) of the whole video clip. We compare the PSNRs, under the same network conditions, of the reconstructed
video sequences at receiver side achieved by using the proposed system to those achieved by using the existing system of non-content-aware analysis, where the video clip is transmitted with fixed QP and AMC scheme. In the experiments, we set QP to 20 and AMC to 3 for the existing system, respectively.

The relation between playback deadline $T_t$ and frame rate $R_{frame}$ meets Eq. 4:

$$T_t = \frac{1}{R_{frame}}$$  \hspace{1cm} (4)

On the basis of this, we consider three different playback deadline values, 20 ms, 30 ms, and 40 ms, respectively. The received video quality achieved through the proposed system in the experiments compared with that through the existing system is demonstrated in Fig. 6. We can observe that in the proposed system, the human motion region based on content-aware analysis has 3-5 dB PSNR improvement over the existing system. Meanwhile, the performance gain of the human motion region over the existing system is even larger in the case of 20 ms than that in the other two cases, indicating that the more stringent the single-packet delay deadline, the more PSNR improvement of the human motion region the proposed scheme can achieve. In other words, the proposed framework is extremely suitable for delay-stringent wireless networks.

To explain this, in a more stringent environment, the packet is more likely to miss the playback deadline. Therefore, more error concealment will be used to play back the video, resulting in degradation of the user-received video quality. By adopting the proposed scheme, the optimized parameters can be chosen dynamically to try to meet this playback deadline. Accordingly, higher PSNR gain can be achieved. However, in a less stringent environment, more packets can meet the delay bound even without using the proposed optimization scheme, so there is less opportunity for the proposed scheme to take effect. Therefore, the proposed joint optimization system can choose the best coding parameters and the optimal AMC scheme to decrease packet loss rate by dynamically adapting to the varying channel quality, while trying to avoid missing the packet playback deadline.

Additionally, in the experiments the content aware analysis helps reduce the amount of video traffic by around 50 percent compared to the existing non-content-aware analysis system.

**Conclusions**

In this article we propose an e-healthcare system based on video content analysis and quality-driven content-aware wireless streaming for remote human gait tracking. The proposed scheme can significantly reduce the reliance on traditional marker-based gait data collection facilities, providing a low-cost high-accuracy gait tracking system. A distortion-delay framework has been proposed to optimize the wireless streaming for delay-bounded retrieval of the collected video data, where key system parameters residing in different network layers are jointly optimized in a holistic way to achieve the best user-perceived video quality over wireless environments. Experimental results have demonstrated that the proposed system can provide great convenience and cost effectiveness for fast prognosis and diagnosis of pathological locomotion biorhythm over resource-constrained wireless networks.
Figure 6: PSNR comparison of different single-packet delay deadlines.
References


Biographies

Haiyan Luo [S’09] received his B.S. degree from Dalian Jiaojong University (formerly Dalian Railway Institute), and his M.S. degree from Dalian University of Technology, China, He is currently pursuing his Ph.D. degree in the Computer and Electronics Engineering Department of the University of Nebraska-Lincoln. His research interests focus on networking and distributed systems. Previously, he worked for Bell Laboratories, Lucent Technologies as a member of technical staff. He has served as a referee for IEEE Transactions on Wireless Communications, IEEE Transactions on Vehicular Technology, and several conferences such as IEEE INFOCOM, IEEE GLOBECOM, and IEEE ICC.

SONG CI [S’98, M’02, SM’06] (sci@engr.unl.edu) is an associate professor in the Computer and Electronics Engineering Department of the University of Nebraska-Lincoln. He is director of the Intelligent Ubiquitous Computing Laboratory (iUbi-Comp Lab) and holds a courtesy appointment of UNL Ph.D. in the Biomedical Engineering Program. He is also affiliated with the Nebraska Biomechanics Core Facility at the University of Nebraska at Omaha and the Center for Advanced Surgical Technology (CAST) at the University of Nebraska Medical Center, Omaha. His research interests include dynamic complex system modeling and optimization, green computing and power management, dynamically reconfigurable embedded systems, content-aware quality-driven cross-layer optimized multimedia over wireless, cognitive network management and service-oriented architecture, and cyber-enable e-healthcare. He is an Associate Editor of IEEE Transactions on Vehicular Technology and serves as a Guest Editor of IEEE Transactions on Multimedia (Special Issue on Quality-Driven Cross-Layer Design for Multimedia Communications); Guest Editor of IEEE Network (Special Issue on Wireless Mesh Networks: Applications, Architectures, and Protocols); and Associate Editor of four other international journals. He served as TPC co-chair of IEEE ICCCN 2009 and TPC member for numerous conferences. He is Vice Chair of the Computer Society of the IEEE Nebraska Section and is a member of the ACM and the ASH RAE. He is the SIG chair of the Quality-Driven Cross-Layer Design of Multimedia Communications Technical Committee (MMTC) of the IEEE Communications Society.

DALEI Wu [S’05] received his B.S. and M.Eng. degrees in electrical engineering from Shandong University, Jinan, China, in 2001 and 2004, respectively. He is currently pursuing a Ph.D. degree in the Department of Computer and Electronics Engineering of the University of Nebraska-Lincoln. His research interests include cross-layer design and optimization over wireless networks, wireless multimedia communications, large-scale system modeling and analysis, and approximate dynamic programming.

NICHOLAS STERGIOU is a Professor at the University of Nebraska at Omaha. He is the Director of the Nebraska Biomechanics Core Facility and also holds an appointment of Professor in the Department of Environmental, Agricultural and Occupational Health Sciences at the University of Nebraska Medical Center. His research interests include nonlinear dynamics, variability and coordination of human movements, lower extremity function, infant postural development, and musculoskeletal injuries. He serves as a reviewer for numerous professional organizations, including the National Institutes of Health,
the National Science Foundation, and several professional journals. He also holds several courtesy appointments including the Department of Physical Therapy at Creighton University, the Departments of Pediatrics, and the Department of Surgery at the University of Nebraska Medical Center, and the Psychology Department at the University of Nebraska at Omaha.

KA-CHUN S1u is an Assistant Professor at the College of Public Health, University of Medical Center (UNMC), Omaha, NE. He is also an Adjunct Assistant Professor at the College of Education, University of Nebraska at Omaha and at the Department of Surgery, UNMC, Omaha, NE. His research interests include balance control and gait in aging, motor relearning and rehabilitation in the elderly with balance problems. He is also a Physical Therapist emphasizing on Neurological pathologies and motor disorders.