Appl. Math. Inf. Sci. **9**, No. 6, 3263-3279 (2015) 3263

Applied Mathematics & Information Sciences
An International Journal

http://dx.doi.org/10.12785/amis/090654

A Bandwidth Efficient Pareto Minimal Approach for Gesture based Video Streaming

S. Swarna Parvathi[1,*], *K. S. Easwarakumar*[2], *N. Devi*[1] *and Raju Das*[3]

[1] Department of Information Technology, Sri Venkateswara College of Engineering, Pennalur, India
[2] Anna University- CEG Campus, Guindy, Chennai, India
[3] Space Applications Centre, ISRO, Ahmadabad, India

Received: 15 Mar. 2015, Revised: 13 May 2015, Accepted: 14 May 2015
Published online: 1 Nov. 2015

Abstract: Deaf, Hard-Of-Hearing and Speech-Impaired (D-HOH-SI) individuals have a specific interest in the development of affordable high-quality videoconferencing as a means of communicating with their family members and peers using sign language. Unlike Video Relay Service, which is intended to support communication between a caller using sign language and another party using spoken language, videoconferencing can directly be used either between two deaf signers or between a caller using sign language and the other using spoken language without the need of an interpreter. This paper proposes a Bandwidth Aware Gesture Based Layered (BAGBL) framework for sign language recognition based video conferencing application. Assuming a D-HOH-SI individual at the sender side, the proposed framework uses shape energy trajectory of hand sign gesture for video layering, a Multi-dimensional Multiple-choice Knapsack Problem (MMKP) based gradational hull pareto minimization heuristic called MMKP based Pareto Minimization Heuristic for Substream Scaling (MPMHSS) and a heuristic for substream scheduling which is based on Dynamic Multilevel Priority (DMP) called Modified DMP packet scheduling (MDMP) mechanism. At the receiving side, our framework includes an automatic sign language recognizer to recognize the sign language gesture and a speech synthesizer to convert the recognized words to speech. Our framework intelligently forms and selects video layers from a video sequence to maximize the video quality. Using extensive simulation and mathematical analysis we show that the proposed solution: (i) is efficient in terms of recognition rate (ii) achieves high radio resource utilization, (iii) maximizes the received video quality.

Keywords: Video Conferencing, Sign Language Recognition, Multimedia Streaming, BAGBL, MPMHSS, MDMP, Key Frame Extraction

1 Introduction

Video communication or in particular video conferencing is efficiently used for interpersonal communication [1] such as as business meetings, social communication or e-learning. This technology supports real-time communication between users separated by a distance. For such communication, network connectivity is an important aspect so as to transmit video and audio signals over distances.

The major strengths of video conferencing technology is the opportunity it offers to people with disabilities to get a face-to-face experience with each other, with service providers and business associates even if they are located across town or around the world. It is a boon to sign language users can communicate in their own language (meaning their own cultural expressions and dynamics) among themselves and with others who are at a distance. There is a necessity for sufficient video quality to capture the nuances of these sign languages rapidly changing hand and finger movements. But, networks suffer from bandwidth limitation which can cause packet loss, jitter and delay [2].Also, if the transmission bit rate exceeds the bandwidth limitations, there is a possibility of poor video experience which results in difficulty of interpretation of sign languages.This results in stringent QoS demands and hence video conferencing over the networks becomes a real challenge.

Advancements in hierarchical coding schemes such as the H.264/SVC [3] have proved to be efficient for bandwidth limited transmission of video over wireless networks.This concept can be extended to gesture based

* Corresponding author e-mail: swarnamurugan2002@gmail.com

video streaming as well. We propose a layered approach to put the data of different importance into different layers. With such features this layered approach is more suitable when the video is transmitted over an error prone channel with fluctuating bandwidth. Specifically, this involves generating a (bit stream) coded representation by decoding of appropriate subsets to reconstruct complete pictures of resolution or quality commensurate with the portion of the bit stream decoded.The base layer is formed by decoding the minimum stream subset (key frames) and the enhancement layer(s) are formed by decoding the remaining bits in the bit stream which contribute to a richer viewing experience due to a higher resolution or quality as compared to the base layer.We experimentally prove that, the recognition rate of sign gestures using this layered approach is comparable to the non-layered one.

To summarize, we propose a layered approach, where the captured video is to be separated into multiple layers of resolutions, frame rates, and quality for streaming. We use the term streaming and video conferencing interchangeably throughout the paper.

Specifically, the contributions of this paper are twofold:

1. The formation of hierarchical layers using extraction of key frames.Key frames are extracted by finding the minimal set of key frames which cover all significant events based on the curvature of the ”shape energy” trajectory.

2. The formulation and implementation of a Multi-dimensional Multiple-choice Knapsack problem (MMKP) based heuristic called A MMKP based Pareto Minimization Heuristic for Substream Scaling for substream scaling (MPMHSS) by using Pareto minimization concepts and a substream scheduling heuristic called MDMP, Modified Dynamic Multilevel Priority packet scheduling mechanism based on Dynamic Multilevel Priority (DMP) [4] to schedule the selected substreams through the wireless network.

Remainder of the paper is organized as follows: Section 2 provides a brief review of the related work. The proposed Bandwidth Aware Gesture Based Layered (BAGBL) framework for streaming gesture based video among multiple wireless receivers is discussed in Section 3. We state our problem and present the analytical formulation for the same in Section 4 stating the heuristics MPMHSS and MDMP. The time complexity of our heuristics is evaluated in Section 5. Section 6 presents the simulation set up for evaluating our framework analyzing various metrics on gesture based test data sets. Section 7 concludes the paper.

2 Related work

2.1 Sign Language Recognition

Sign language is a language which exploits unique features of the visual medium through spatial grammar .It relies on sign patterns viz., the movements and orientation of the arm and the body language of the person to facilitate understanding between people in which communication is viable without the means of acoustic sounds.Attempts to automatically recognize sign language began to appear in the literature in the 90s.

Various approaches to handle the data acquisition problem towards automated SL recognition have been proposed, that employ sensing devices, which include cameras [5,6] to wearable tracking devices [7], i.e.,instrumented gloves, body suits, and marker-based optical tracking, or a combination of devices [8,9], for communication between a human and the environment.There is variation in each sensing technology in several dimensions, which include accuracy, range of motion, resolution, , user comfort, latency, and cost [10]. Glove based gestural interfaces characteristically require the user to wear a cumbersome device and hold back the ease and naturalness of the users interaction with the computer. Vision-based techniques, while overcoming this, need to compete with other problems related to, in some cases, specific environment limitations [10]. An alternative SL information acquisition approach was lately introduced, referring to the use of surface electromyogram (sEMG) (acquired from the users forearm), either exclusively [11,12,7,13] or combined with data from a triaxial accelerometer (3D-Acc) (placed on the users wrist) [11,14] with the use of a portable Bluetooth-enabled device. sEMG signals were acquired from the skin above the muscles involved during gesturing with noninvasive surface electrodes.

The mathematical tools involved in automated SL recognition solutions include approaches involving statistical modeling, computer vision, pattern recognition, and image processing.SushmitaMitra [15], provides a survey on gesture recognition, with particular emphasis on hand gestures and facial expressions. The major tools surveyed for this purpose include HMMs, particle filtering and condensation algorithm, FSMs, hybridization of HMMs and FSMs, Soft computing tools, fuzzy logic and ANNs. A sign language recognition system using Kinect [16] was developed by researchers from Microsoft Research Asia and the Chinese Academy of Sciences which provides two modes viz., a Translation Mode, which translates sign language (word or sentence) into speech and a Communication Mode in which a person can orally communicate with the signer through an avatar.But this system is confined to Chinese sign language.

Li-Chun et.al., [17] developed a model for measuring the similarities between two sign language videos. The vision element of the model is distance based on Volume

Local Binary Patterns (VLBP), which is robust for motion and illumination. Semantic component of the model computes semantic distance based on definition of sign language semantic, which is defined as hand shape, location, orientation and actions. While quantizing the sign language semantic, contour is used to measure shape and orientation; trajectory is used for measuring location and movement. Jessica [18] has developed a better theoretical model, the Human Signal Intelligibility (HSI), to evaluate intelligibility of lowered video quality for the purpose of reducing bandwidth consumption and extending cell phone battery duration.The goal is to advance mobile sign language video communication so that it does not rely on higher cellular network bandwidth capacities.

A Region of Interest-based Adaptive Scheme (ROIAS) in which the adaptation of streamed multimedia content is different in the regions within each frame based on the users interest in them was presented by Bogdan et al. [19].

2.2 Adaptive Multimedia Streaming

There are many performance issues involved while delivering multimedia over variously loaded best-effort networks to heterogeneous users in terms of connectivity, device characteristics and expectations. Among the most significant causes of degradation of performance when streaming multimedia are the low bandwidth available, lossy connections, highly variable background traffic and highly loaded delivery conditions. The combined effect of these parameters ultimately affects end-user Quality of Experience (QoE). As QoE is difficult to assess, research has instead focused on proposing techniques to increase the Quality of Service (QoS) level.

Various solutions ranging from bandwidth over-provisioning to traffic engineering were proposed which were either very expensive, difficult to deploy or provided limited scalability and flexibility.Among the most successful solutions are the adaptive multimedia streaming schemes which adjust the bandwidth used by the applications by increasing or decreasing both the transmission and content encoding rates according to the existing network conditions. Among the approaches proposed, network or transport level adaptive solutions such as TFRCP [20], LDA+ [21] and RAP [22] provided certain level of QoS when streaming multimedia over wired networks, but were poorly linked to end-user perceived quality. Various solutions were proposed for adaptive multimedia transmissions over wireless access networks [23] or wireless ad-hoc networks [24]. Among the proposed solutions are adaptation mechanisms at the level of layers [23] or objects [25], fine-granular scalability schemes [26] and perception-based approaches [27]. Complementing these approaches the IEEE 802.11e standard [28] provides QoS features that may help improving users QoE allowing for multiple-priority-based distribution of multimedia content. [29] proposed a TCP-friendly video transport protocol targeting for wireless environment, but it is still content-agnostic. [30, 31, 32, 33] rely on using a video rate-distortion model to solve network resource allocation while providing video content-awareness. The model is however only applicable for videos at a fixed frame rate. Cross-layer design is also investigated in the literature to improve the video adaptation over wireless [34, 35, 36, 37] that get closer to the users and try to achieve higher end-user perceived quality of multimedia streaming. However only the Quality Oriented Adaptation Scheme (QOAS) [38] involves a user perceived quality estimation in the feedback-based multimedia adaptation process. Hence, when used for adaptively streaming multimedia in both wired and wireless environments, QOAS shows appreciable improvements in end-user QoE. In a recent work, [39], proxy-assisted video adaptation is considered for the case where multiple streams share a common backbone network and the video rate-distortion model is used. Most recently, [40] developed a novel analytical rate-quality model, that relates the maximum notable quality for a given video rate, for scalable video with both temporal and amplitude scalability. They came up with an iterative multi-stream rate allocation scheme at the proxy that can maximize a weighted sum of received video quality at all receivers, given the link bandwidths of all receivers.

All the research presented above are oriented specifically either towards sign language recognition systems or adaptive multimedia streaming. The concept of bandwidth proficient transmission of sign language data over networks is lacking on a large scale. To address the above issue, we propose a BAGBL framework as a novel bit-rate adaptation technique for gesture based video streaming by forming hierarchical layers using the shape energy trajectory of sign gestures.

3 Bandwidth Aware Gesture Based Layered (BAGBL) Framework and Algorithms for Layer Selection

In this section, we develop a Bandwidth Aware Gesture Based Layered (BAGBL) framework for streaming gesture based video among multiple wireless receivers. The framework is defined for gesture based video conferencing and can be easily adapted to accommodate other multimodal based streaming applications.

The BAGBL framework is designed to facilitate hand sign gesture based video conferencing over wireless networks with bandwidth fluctuations. The system is composed of three major subsystems: Key Frame Extraction (KFES) Subsystem, Layering Subsystem (LS) and Streaming Subsystem (SS) as shown in Fig. 1. Video analytics applications [41] that suffer from the problem of processing large number of video frames use Key frame extraction as a pre-processing step. This concept can be

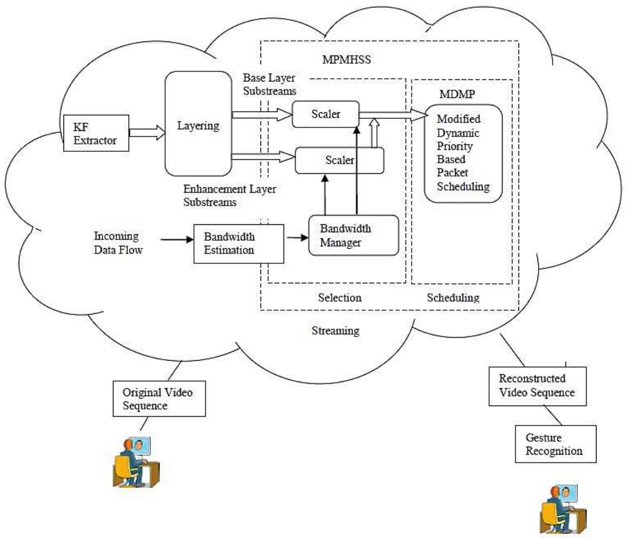

Fig. 1: The BAGBL Framework

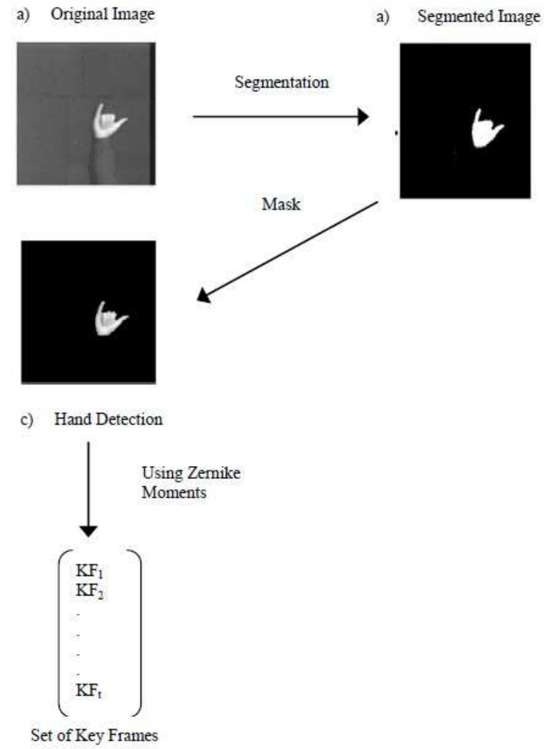

Fig. 2: Key Frame Extraction

applied for video streaming applications as well. Key frame extraction, is the problem of finding the minimal set of key frames that cover all significant events or maximize the number of key frames while minimizing redundancy of information in these key frames. Video layers are generated by arranging the video frames according to the amount of information present. Video streaming is achieved by just dropping the least important levels (layers) depending on the available bandwidth and client requirements.

3.1 Key Frame Extraction (KFES) Subsystem:

3.1.1 Pre Processing:

For extraction of candidate frames from the sequence we perform certain pre-processing. Pre processing involves segmentation, masking and hand detection. The first stage is segmenting the hand. Many methods have been proposed for hand segmentation, we use skin color to detect and segment hands [42], but unfortunately by itself is not a reliable modality. It is assumed to have a uniform background and clothes for simplification. Thus, the segmentation can be easily performed using a threshold. After that, the segmented hand (binary image) could be used as a mask, for extracting the Zernike moments as shown in Fig. 2.

3.1.2 Key frame extraction:

To select only those frames that are able to describe adequately the performed gestures [43], key frames are extracted based on the curvature of the “shape energy” trajectory to create a content based representation for them.

Let V_t denote the original video sequence at an instant t captured at a rate of N_f frames per second represented by a set of frames $F(t)$ denoted as $V_t = F(t)$.The aim is to derive a captured set of key frames $KF_t = F_k(t)$ from V_t such that the gesture energy J (as shown in equation 1) represented by the sum of the squared coefficients of the Zernike moments is maximum.

$$J = \sum_{n=0}^{n_{max}} \sum_{m} \|Anm\|^2. \tag{1}$$

Where n_{max} is the selected order of the moments. n_{max} is selected as 2, the second derivative of J denoted by J'' which gives the curvature measure. Local minima indicate time instances of low shape variation whereas local maxima correspond to peak shape variation.

Let X_M contain the time instances of frames corresponding to the local maxima of J'' estimated as in equation 2 & 3

$$X_M = F(t) : J''(F(t)-1) < J''(F(t))\&J''(F(t)) > J''(F(t)+1) \tag{2}$$

and X_m contain the time instances of local minima of J' estimated as

$$X_m = F(t) : J''(F(t)-1) > J''(F(t))\&J''(F(t)) > (F(t)+1) \quad (3)$$

X_M that contains the local maxima is denoted as the set of key frames extracted from the video sequence, $KF_t = X_M$.

3.2 Layering Subsystem:

A video sequence may be represented using an essential base layer (also called the main profile) and one or more optional enhancement layers (called scalability profiles). The base layer constructs the coarse or base representation of the stream, and the enhancement layers successively improve it.

The base layer is formed by the second order moments of the extracted key-frames as described in Section 3.1 which preserve the semantics of the gesture. The higher order moments contribute to the enhancement layers. At the client side, the image can be reconstructed using a set of moments through order M as given by equation 4

$$\hat{f}(x,y) = \sum_{n=0}^{n_{max}} \sum_{m} A_{nm} V_{nm}(\rho,\theta) \quad (4)$$

where $\hat{f}(x;y)$ is the reconstruction of the original image $f(x;y)$, n is a positive integer, m is an integer such that $|m| <= n$, and $(n-|m|)/2 = 0$, and n_{max} is the maximum value of n (maximum order). As higher values of n_{max} capture a lot of noise, n_{max} is determined by computing moments up to an order where there is only a 1% difference between the reconstruction and the original.

3.3 Streaming Subsystem:

Streaming refers to the transmission of different video layers in different bit streams called sub streams. Less bandwidth is needed for transmission of the base layer due to its coarser quality; while the enhancement layers require more bandwidth due to finer quality.

4 Problem Formulation

We formulate our problem of gesture based video streaming over wireless networks as two subproblems viz., Substream Scaling by indicating a MMKP based approach which uses pareto minimization techniques and Substream Scheduling which is based on Dynamic Multilevel Priority [4]. We devise our own heuristic to solve the above mentioned MMKP based video streaming procedure. The selected substreams are scheduled through the wireless network using our own procedure called MDMP, Modified DMP packet scheduling mechanism.

4.1 Substream Scaling:

The original video sequence is encoded into S scalable streams. Each scalable stream "s", $1 \le s \le S$ has a number of layers (a base layer and one or more enhancement layers) as described in the previous section. The objective of the proposed MMKP based video streaming is to choose l sub streams, where a sub stream l includes layer L and all layers below it such that the total average quality across the video streams is maximized subject to the client-side resource constraints. Each sub stream has an associated Bit Rate (b_{sl}), Quality (PSNR) (q_{sl}) and Shape Energy (e_{sl}). Let the amount of available resources be R = $(R_1, R_2, \ldots, R_m)$. As a simple case, we choose two resource constraints namely R_1 = maximum capacity of network bandwidth, denoted by B and R_2 = maximum shape energy of hand sign gestures, denoted by E that can be sought from the hand sign gesture database which contains the shape energy of each and every gesture. The same is available both on the sender as well as the receiver side. At the sender side, this is used for layering purpose and on the receiving side this is used for recognizing the hand gesture.

We show that the problem of selecting the sub streams is equivalent to the NP-hard Multi-dimensional Multiple-choice Knapsack Problem.

The Multidimensional Multiple-choice Knapsack Problem (MMKP) is a variant of the classical $0-1$ KP as indicated: Let there be n groups of items. Group i has l_i items. Each item of the group has a particular value and it requires m resources. The objective of MMKP is to pick exactly one item from each group for maximum total value of the collected items, subject to m resource constraints of the knapsack. In mathematical notation, let v_{ij} be the value of the j^{th} item of the i^{th} group, $\overline{r_{ij}} = (r_{ij1}, r_{ij2}, \ldots, r_{ijm})$ be the required resource vector for the j^{th} item of the i^{th} group, and $\overline{R} = (R_1, R_2, \ldots, R_m)$ be the resource bound of the knapsack.

Formally, MMKP is expressed as follows:(specified in equations 5 through 13)

V_{max} = Maximize,

$$\sum_{i=1}^{n} \sum_{j=1}^{l_i} x_{ij} v_{ij}; i = 1,2,\ldots,n; j = 1,2,\ldots,l_i \quad (5)$$

such that

$$\sum_{i=1}^{n} \sum_{j=1}^{l_i} x_{ij} r_{ijk} \le R_k; k = 1,2,\ldots,m. \quad (6)$$

$$\sum_{i=1}^{n} x_{ij} = 1 \quad (7)$$

$$x_{ij} \in 0,1 \quad (8)$$

Consequently, our problem is to determine

$$Maximize, \sum_{s=1}^{S} \sum_{l=1}^{L} x_{sl} q_{sl}; s = 1,2,\ldots,S; l = 1,2,\ldots,L. \quad (9)$$

Subject to,

$$\sum_{s=1}^{S}\sum_{l=1}^{L} x_{sl} b_{sl} \leq B \tag{10}$$

$$\sum_{s=1}^{S}\sum_{l=1}^{L} x_{sl} e_{sl} \leq E \tag{11}$$

$$\sum_{l=1}^{L} x_{sl} = 1 \tag{12}$$

$$x_{sl} \in 0,1 \tag{13}$$

Here x_{sl} is either 0 or 1. 0 implies a layer l of stream s is not picked and 1 implying layer l of stream s being picked. Table 1 list the symbols used.

Table 1: Table of Symbols

Symbols	Description
S	Number of Streams
s	Scalable Video Streams
L	Number of layers
l	Substream
q_{sl}	PSNR of substream *sl*
b_{sl}	Bit rate of substream *sl*
e_{sl}	Shape Energy of substream *sl*
x_{sl}	0 or 1

Here we present our heuristic for substream scaling over wireless networks.

4.1.1 Concepts:

a) Quantity: A quantity is a parameter, quality metric, or any other quantified aspect. It is represented as a set Q with a partial order $\preccurlyeq Q$ that denotes preferred values (lower values being preferred). The subscript is dropped when clear from the context. If $\preccurlyeq Q$ is total, the quantity is basic. For example, resource costs r_1 and r_2 are taken from finite intervals of non-negative natural numbers that form two basic quantities with the usual total order $\leq$; value v is taken from the basic quantity of the natural numbers IN with total order $\geq$ as the preference.

In our context quantity is the tuple consisting of PSNR (v=Quality), Resource Constraint 1 (r_1=Bandwidth), Resource Constraint 2 (r_2=Shape energy).

b) Configuration Space: A configuration space S is the Cartesian product $Q_1,\ldots,Q_n$ of a finite number of n quantities; a configuration $l = (l_1,\ldots,l_n)$ is an element of such a space, where l Q_k may be used to denote l_k. Groups of layers are captured as configuration sets. The selection of items (substreams) from different groups results in a combined value (PSNR) and resource cost that can be captured by element-wise addition on the configurations representing the items. Such combinations of items are again configurations. These configurations provide partial solution candidates. All possible combinations of configurations from two configuration sets (items or partial solutions) can be captured by taking a tensor product $\otimes$. Tensor product is defined as follows: $L_1 \otimes L_2 = l_1 + l_2 \mid l_1 \in L_1; l_2 \in L_2$, for configuration sets L_1 and L_2, with + denoting the element-wise addition on configurations.

c) Dominance relation: A dominance relation $S_1 \leqslant S_2$ on configuration space S defines preference among configurations. If l_1; $l_2 \in S$, then $l_1 \preccurlyeq l_2$ iff for every quantity Q_k, the k-th component of S,$l_1 Q_k \preccurlyeq Q_k\, l_2(Q_k)$. If $l_1 \preccurlyeq l_2$, l_1 then dominates l_2, expressing that l_1 is in all aspects at least as good as l_2. Dominance is reflexive, i.e., a configuration dominates itself. The irreflexive strict dominance relation is denoted as $\prec$.

d) Pareto points: A configuration is a Pareto point of a configuration set iff it is not strictly dominated by any other configuration. Configuration set L is Pareto minimal iff it contains only Pareto points, i.e., for any l_1; $l_1 \in L$, $l_1 \not\prec l_1$.

4.1.2 A MMKP based Pareto Minimization Heuristic for Substream Scaling (MPMHSS):

We start with finding a feasible solution by devising a heuristic which uses the concepts of Pareto minimization [44], calculating fractional solutions and iteratively combining them to find a final feasible solution. Fractional solutions are found in order to minimize the time and space complexity. Our heuristic approach provides a computational efficiency of $O(\mathrm{M}(max(M_{max}\log M_{max}, M^3)))$.The Pseudo code for MPMHSS is presented in Algorithm 1.

Lines 1 to 3 finds Pareto points in each configuration set, as dominated configurations cannot contribute to an optimal solution. A for-loop applies the Pareto-algebraic min operation. For implementing the above minimization procedure, we slightly modify the Melkman's convex hull algorithm to linear time such that the vertices appear in counter clockwise order around the convex hull at any stage of the algorithm.

Line 4 adds aggregate resource costs to each configuration. Aggregate resource [45] is a projection of the resource vector on current resource usage vector. The main idea is to penalize the not yet picked items depending on the current resource state. The aggregate resource required by item i is computed as shown in equations 14 and 15:

$$a_i = (r_1c_1 + r_2c_2 + \cdots + r_mc_m)/|c| \tag{14}$$

$$= r_i.c/|c| \tag{15}$$

Algorithm 1: Pseudocode for MPMHSS

Input: S, Vector of configuration set, b, Vector of bounds
Output: *result*, a maximum value of configuration which is a solution to the MMKP instance.

1 **begin**
// Keep only Pareto points
2 **forall the** $l_i \in S$ *do min* l_i **do** ;
3 ;
// Manipulate all configurations with aggregate resource usage
4 **forall the** $l_i \in S$ *do Concat(*l_i*, aru)* **do** ;
5 ;
// Initialize rs with first configuration set $S[1] \in S$
6 $rs = S[1] \in S$;
7 $S* = S - S[1]$;
// Iterate over all configuration sets
8 **forall the** $L_x \in S*$ **do**
// Partial solutions to be combined with configurations from L_x
9 $L_{ps} = rs$;
// Select atmost N partial solutions depending on the aggregate resource function and atmost N configurations from L_x
10 $L_{ps} = SelectPS(L_{ps}, N)$;
11 $L_x = SelectPS(L_x, N)$;
// Initialize the resultant configuration set rs
12 $rs = \phi$;
// Iterate over all configurations
13 **forall the** $l_i \in L_{ps}$ **do**
14 **forall the** $l_j \in L_x$ **do**
// Compute the tensor product of L_{ps} and L_x
15 $\bar{l} = l_i \otimes l_j$;
// Compute l' the candidate set of feasible solutions by enforcing the resource bounds
16 If feasible ('l'',$\underline{b}$) then // Minimize the resultant set
// Check whether the current configuration 'l'' is dominated
17 dominated=false // Iterate over all configurations so far in the result set
18 **forall the** $\bar{l'} \in rs$ **do**
// Check if configuration $\bar{l'}$ of resultant set rs is dominated; if so remove it
19 **if** $l' \prec \bar{l'}$ *then* $rs = rs \setminus \{l'\}$ **then**
// Check if configuration l' is itself dominated; if so, stop further minimization
20 **else**
21 If $\bar{l'} \prec l'$ then dominated = true break
22 **end**
23 ;
24 **end**
// Add configuration l' to the result set if not dominated
25 If not dominated then $rs = rs \cup l'$;
26 **end**
27 **end**
28 **end**
29 $result = max(rs)$;
30 Return *result* ;
31 **end**

Where c = $(c_1, c_2, \ldots, c_m)$ is the current resource usage vector, $r_i = (r_1, r_2, \ldots, r_m)$ is the resource requirement of item i.

Lines 6 & 7 does the initialization for the compositional computations. During the compositional steps, a configuration set L_{ps} of partial solutions is maintained. This set is initialized with $S[1]$,the first configuration set in vector S, which is removed from the vector of configurations subsequently.

Lines 8 to 28 implement the compositional steps. It iterates over the list of remaining configuration sets. In each iteration, it performs the following actions.

Line 9: Initializes the set L_{ps} with the resultant set rs which contains the partial solutions obtained from the previous iterations.

Lines 10 & 11: Procedure Select PS selects (atmost) N partial solutions from configuration set L_{ps} and atmost N configurations from configuration set L_x are selected.

Line 12: Clear the resultant set rs for storing the new partial solutions which is computed through lines 13 and 28.

Lines 13 to 28: Iterate over all configurations.

Line 15 takes the tensor product of L_{ps} and L_x.

Line 16 checks feasibility of newly created configurations, enforcing the resource constraints. i.e, the item with maximum value per unit of aggregate resource v_i/a_i among the not yet picked item is picked.

Lines 17 to 28 check whether any compound configurations in the result set that were Pareto points so far are dominated by the new compound configuration , or whether the new configuration is itself dominated. If not dominated, adds the new configuration to the result set.

Lines 29 & 30: rs contains feasible Pareto-optimal configurations that are all solution candidates. Any maximal-value configuration among these candidates is typically already a good solution.

4.2 Substream Scheduling:

Once the substreams are selected, we have to schedule the transmission of packets on the wireless medium according to their substream QoS specifications and priorities. Depending on the channel conditions, a substream might be dropped for a period of time in order to accommodate higher-priority substreams. Thus, a scheduler provides a scaling function as well; however, its scaling function is a result of its scheduling function.

Multicast Broadcast Services (MBS) is defined as a service available in broadband wireless networks, when multiple Mobile Stations (MS) connected to a Base Station (BS) receive the same information or when multiple BSs transmit the same information. In fact, this allows in resource saving as a single radio pipe may be allocated for all users registered to the same service instead of allocating as many pipes as there are users.This is of particular interest for video conferencing applications where, at the same time, several users under the same coverage area are connected to the same service. Each selected substream is to be broadcast using MBS to a group of MS. At the BS, the MBS module allocates a fixed-size data area in the download section of each TDD frame. A scheduling window is composed of a number of MBS data areas, the data rates can be assumed to be constant during a scheduling window, but varying across scheduling windows. Data from the video streams are to be allocated to the MBS areas in the scheduling window. Due to the variable bit rate (VBR) nature of the video streams, the aggregate data rates may exceed the broadcast channel capacity. Hence, in each scheduling window, we need to decide which layers to send for each stream for which we devise a Modified Dynamic Multilevel Priority packet scheduling heuristic subject to the following constraints.,

1. Any selected sub stream l, which includes layer L and all layers below it, must fit into the MBS area in the current scheduling window.

2. The base layer of each stream must be transmitted to guarantee a basic service level agreement.

4.2.1 Modified Dynamic Multilevel Priority packet scheduling:

In wireless networks, especially for real time applications, efforts to reduce end to end transmission delay must be considered. Though various packet scheduling algorithms are available [46,47] it is more important to assure the delivery of various types of packets depending upon the priority. We therefore propose a modified version of the Dynamic Multilevel Priority (DMP) packet scheduling [4] called MDMP (Modified Dynamic Multilevel Priority) packet scheduling. The pseudocode for the same is presented in Algorithm 2.

In MDMP, two types of priority queues are maintained. Base layer streams are placed at the highest priority (priority 0) queue, enhancement layer streams are placed in the subsequent priority '1' to priority 'n' queues as in lines 4-15. It should be noted that at any instance, all the queues should contain a base layer and its corresponding enhancement layers. To enforce this, a sliding window is assumed which is formed by one base layer (L_0) and its corresponding enhancement layers (L_i) as indicated in line 3. After allocating the streams to the MBS areas from the first subsequence containing the base layer (L_0) and its corresponding enhancement layers (L_i) from the window, subsequently, the window slides to process the next subsequence as indicated in lines 2-3.

Thus, scheduling has two phases: (i) allocating layers among different queues based on the priority of the layers, (ii) allocating streams to MBS areas.All the packets are sent in a non-preemptive fashion and the current available bandwidth is assumed to be greater than the data rate of the base layer packets and the enhancement layer packets are sent till enough bandwidth is available as in lines 16-24.

Algorithm 2: Pseudocode for MDMP Packet Scheduling

Input: A sliding window (SW) formed by one base layer (L_0) and its corresponding enhancement layers (L_i),
$Q[pr]$, the priority queue,
D, the data rate of the MBS data area,
n, the number of enhancement layers,
r_0, data rate of the base layer,
r_k, the data rate of the enhancement layer k, V_t, the video sequence
Output: Packets sent through the network

1 **begin**
2 **forall the** $L \in V_t$ **do**
// Repeat for all subsequences of a video
3 $SW = L$; // Clear the queue
4 **forall the** $j \leftarrow 0$ *to* n **do** ;
5 ;
6 $Q[pr_j] = null$; // Queue the layers according to priorities
7 **forall the** $i \leftarrow 0$ *to* n **do**
8 **if** *type*(L_i) = *Base layer* **then**
9 place L_i in $Q[pr_i]$;
// L_0 is placed in $Q[pr_0]$
10 **else**
11 if type (L_i) = Enhancement layer i then
12 place L_i in $Q[pr_i]$
13 **end**
14 ;
15 **end**
16 Allocate $Q[pr_1]$ to MBS data area;
// Send the base layer
17 $D = D - r_0$;
// Calculate the remaining bandwidth
// Transmit enhancement layers till enough bandwidth is available
18 **for** $k \leftarrow 1$ *to* n **do**
19 select packet from pr_k queue;
20 **if** $r_k < D$ **then**
21 allocate packet to MBS data area;
22 $D = D - r_k$;
23 **else**
24 break
25 **end**
26 ;
27 **end**
28 **end**
29 **end**

5 Complexity Analysis:

5.1 Run-time complexity of MPMHSS:

The running time of MPMHSS is typically dominated by the compositional part of the heuristic. Therefore we limit our analysis to the compositional part of MPMHSS.The first part up to line 6 of the heuristic is ignored as this does not really contribute in any significant way to the run- time.

Assume we have M configuration sets in input vector S and that M_{max} is the maximum number of configurations in any of the sets in S. The complexity of each step in the compositional part is dependent on the number of configurations in the considered configuration sets (L_{ps} and L_x) as well as on the parameter N.

The size of this configuration set is $M = |L_1 \times L_2|$; the worst-case size of the result set is also M^2 (when all compound configurations are Pareto points).The worst-case complexity of lines 13 to 28 is therefore $O(M)$, i.e., linear in terms of the size of the set to be minimized, which is exactly the complexity of the modified Melkman's convex hull [48]. Note that the size of any of the intermediate L_{ps} results is bounded by N^2, because in the worst case all M configurations in the product computed in lines 13 to 28 might be Pareto-optimal.

The two executions of Select PS in every iteration of the loop in lines 10 & 11 of MPMHSS each require a sorting of the configuration set and a rehearse of the result to select M configurations. Therefore, given that the size of L_{ps} may be M and that the size of L_x is at most M_{max}, the complexity of these two reductions is $O(MlogM)$ and $O(M_{max}logM_{max})$.

Since the tensor product computed by the line 15 is $O(M^2)$, the complexity of the minimization procedure (lines 18-24) is $O(M)$,the overall complexity of lines 13 to 28 is $O(M^3)$.

Since $M-1$ compositional steps are involved in MPMHSS, the compositional part has an overall complexity of $O(\mathrm{M}(max(M_{max}\log M_{max},M^3)))$ or $O(M^4)$.

5.2 Run-time complexity of MDMP:

Lines 3-6 perform only initialization and therefore analysis is limited to the compositional part as this half has no significance to the run-time. The complexity for queuing the layers (Lines 7-15) according to priorities is $O(n)$ where n is the number of enhancement layers. The worst case complexity for transmitting the enhancement layers assuming all the enhancement layers can be fit into the available bandwidth is $O(n)$.Therefore the overall complexity is $O(n)$ linear time for a video subsequence.

5.3 Overall complexity of the system:

Combining both MPMHSS and MDMP, the overall complexity of the system is given by equation 16 for a particular subsequence.

$$O(M^4) + O(n) \tag{16}$$

6 Results and Discussion:

This section provides the details of our simulation setup and results. We use this setup to evaluate the effectiveness of our proposed heuristics (MPMHSS and MDMP).

6.1 Simulation Setup

A video conferencing setup is implemented over our own point-to-point WiMAX multimedia broadcast simulator based on Sharangi et al., [49] and evaluated our heuristics in it using actual layered video traces. We developed our own Sign language recognition system based on Qutaishat Munib et al. [50] and a speech synthesizer using TTS (Text to speech software) as part of the video conferencing system to evaluate our framework.

For generating the video traffic we use sign gesture video files from the help of gesture database [51]. This video database consists of 800 videos. Each video represents a gesture of the American Sign Language (ASL).

Fig. 3, presents one in every 5 frames of the "against" word. Fig. 4 shows the second derivative of the shape energy and Fig. 5. shows the respective key-frames.The salient parts of the shots i.e., the 2,7,11,14,24,32,37,39,56,57 frame indices which are detected as key-frames,form the base layer.The enhancement layers are formed by grouping the remaining frames with respect to the above key-frames. For each group, the two frames which have the highest shape energy are selected for further decomposition. This process is repeated until the required number of enhancement layers is attained. It is therefore evident that the frame transmission is by all means adaptable to the network channel capacity and the variations in the characteristics of the same.

The layered video traffic is now streamed through the WiMAX broadcast simulator. The simulation parameters are specified in Table 2.

The reconstructed video at the receiving end is passed through the sign language recognition system which uses filtering, skin color detection, and hand-postures-contours comparison algorithms for detecting face and hand regions, and a systematic process of principal component analysis along with Sober edge detection for identifying the hand gestures. To recognize the sign gestures, Hough transform and Feed Forward back propagation network is used. Results of the system are shown in Fig. 6. And

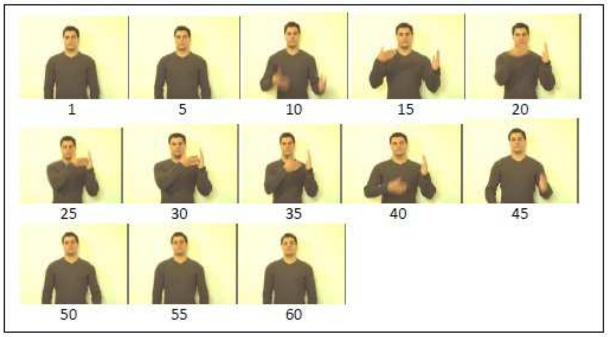

Fig. 3: Frames of the Video Sequence “Against” at intervals of 5

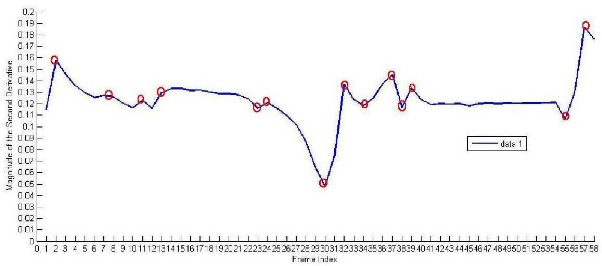

Fig. 4: Second Derivative Zernike Moments of the Video Sequence “Against”

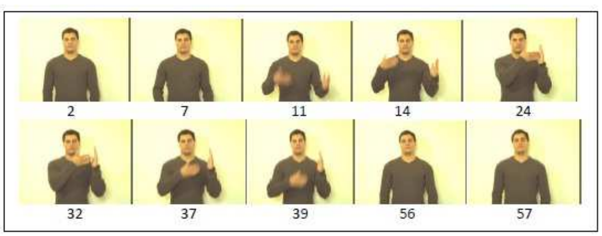

Fig. 5: Key Frames Identified by Finding the Local Maxima of the Zernike Moment

finally we use the Verbose Text to speech software to convert the resultant text to speech.

6.2 Performance Analysis:

The performance of the system is analyzed under various circumstances with respect to the video quality using two different metrics viz., PSNR (Peak Signal to Noise Ratio) and Recognition Rate. We also evaluate the resource utilization of our heuristics in terms of the scheduling window capacity used.

Table 2: WiMAX Simulation parameters

Parameter	**Values**
Modulation Scheme	16-QAM
Coding Scheme	3/4 convolution turbo coding
Channel	10MHz TDD channel
TDD frame size	5 ms
No. of TDD frames	200
MBS data area	50 KB
Broadcast channel bandwidth	10 Mbps

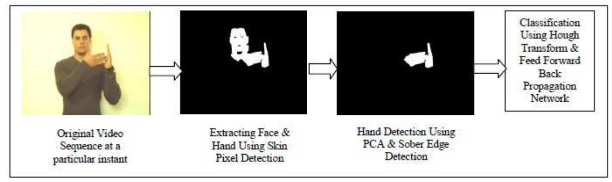

Fig. 6: Output of Sign Language Recognition System for the video “against”

6.2.1 Video Quality with respect to PSNR:

Several objective image and distortion/quality metrics have been identified in literature [52].We choose PSNR as the objective metric despite the fact that it only provides an approximate measure of the quality as subjectively perceived by human observers [53],because, it is simple to calculate and has clear physical meanings. PSNR is mathematically expressed as specified in equations 17 and 18:

$$PSNR = 20log_{10}(MAX_I/\sqrt{MSE}) \qquad (17)$$

where,

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j) - K(i,j)]^2 \qquad (18)$$

MAX_I is the maximum possible pixel value of the image, e.g., when the pixels are represented using 8 bits per sample, MAX_I is 255.

We analyze the performance of the framework in terms of video quality in two different experiments, one by varying the number of streams and the other by the scheduling window capacity.

We measure the video quality of the system by computing PSNR of the streamed video over the network after application of our heuristics for both experiments. Figures 7 to 11 show the results of the PSNR computed for a fixed scheduling window size of 1 sec with the number of streams varying from 10 to 50 for the against, cake, explore, role and table video sequences respectively. We restrict the number of scheduling instances to 100. Fig. 12 to 16 show the individual video qualities for the against, cake, explore, role and table video sequences respectively for a fixed number of streams of 20 and a varying scheduling window capacity of 1 to 10 sec.Results of both the experiments indicate that the proposed heuristics scales well when the number of streams increase and the quality improves as the scheduling window grows.

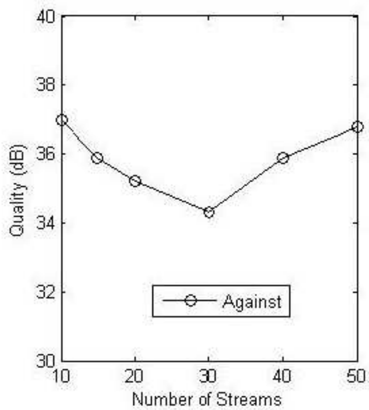

Fig. 7: Video Quality Analysis-Against

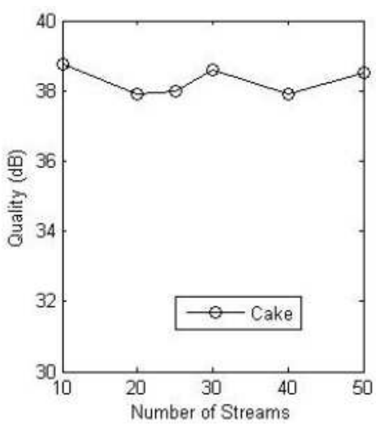

Fig. 8: Video Quality Analysis-Cake

6.2.2 Video Quality with respect to Gesture Recognition:

Several objective and subjective metrics have been identified for gesture recognition. Here we adopt the commonly used Recognition rate and the Likert Score. Recognition Rate provides an objective measure of the video quality whereas Likert score is a subjective metric. The Likert scoring process is performed by choosing 25 randomly-selected persons, and responses are based on a

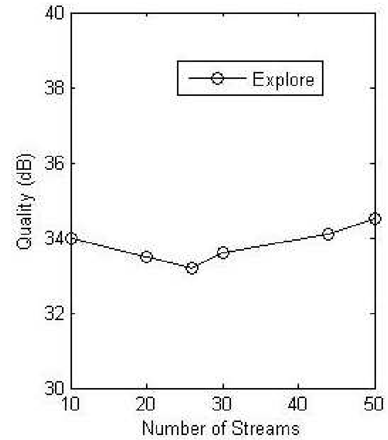

Fig. 9: Video Quality Analysis-Explore

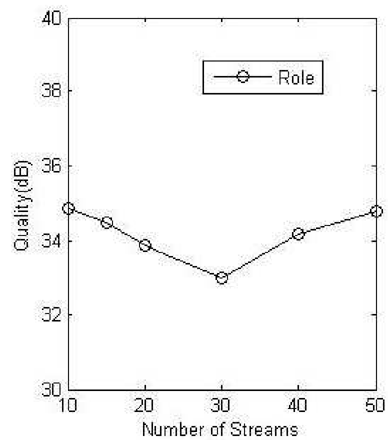

Fig. 10: Video Quality Analysis-Role

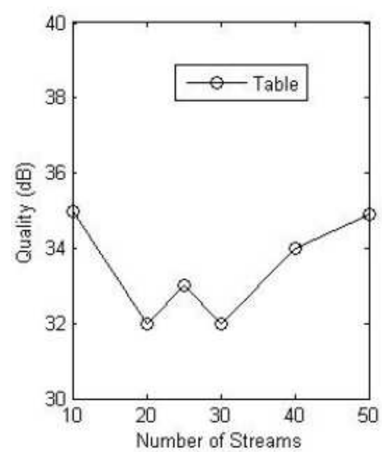

Fig. 11: Video Quality Analysis-Table

7-point Likert scales ranging from 1 -meaning strongly

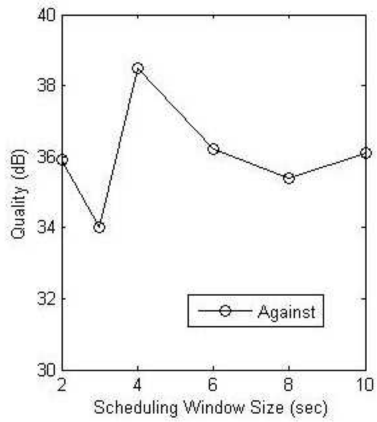

Fig. 12: Video Quality Analysis-Against

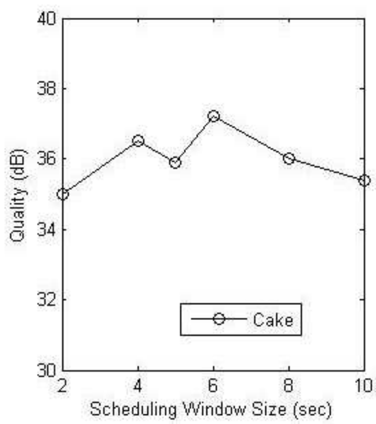

Fig. 13: Video Quality Analysis-Cake

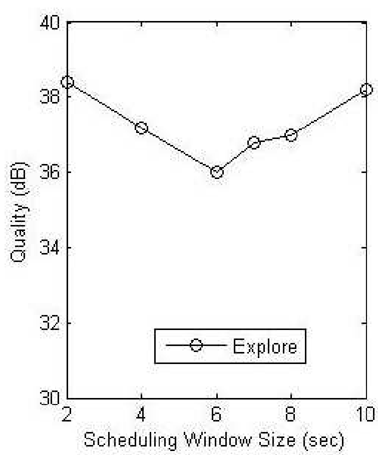

Fig. 14: Video Quality Analysis-Explore

disagree to 7-meaning strongly agree. The valuing is performed considering visual and action effects.For each

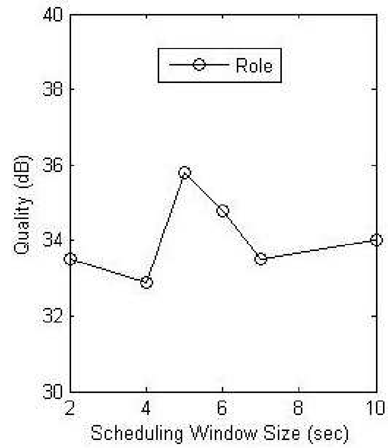

Fig. 15: Video Quality Analysis-Role

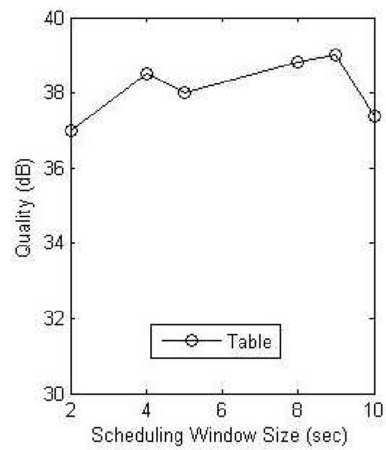

Fig. 16: Video Quality Analysis-Table

video sequence, the mean score of all 25 subjects is projected into a range [0, 1]. Despite it is well known that Recognition Rate only provides an approximate measure of the quality as subjectively perceived by human observers, it is widely used because it is simple to calculate and has clear physical meanings, which is defined as the ratio of the number of correctly classified samples to the total number of samples as shown in equation 19,

$$\text{Recognition Rate} = \frac{\text{Number of correctly classified signs}}{\text{Total number of signs}} * 100 \tag{19}$$

The performance of our system was evaluated by testing its ability to classify signs for both the sender side and the receiver side as shown in Tables 3 and 4. At the sender side recognition of the sign gesture is performed on the original video and at the receiver side, using the reconstructed video. Table 3 indicates the results of the subjective and objective assessment of the captured video and Table 4 that of the reconstructed video after application of our heuristics. Results indicate that the reconstructed video provides a satisfactory recognition rate.

Table 3: Objective and Subjective Assessment Scores of Original Video

S.No	Videos	Objective Assessment (Recognition Rate)	Subjective Assessment (Likert Score)
1	Against	0.8923	0.884
2	Assume	0.8726	0.847
3	Aim	0.8853	0.906
4	Balloon	0.8855	0.903
5	Cake	0.8375	0.895
6	Chair	0.8772	0.884
7	Computer	0.88	0.957
8	Concentrate	0.8792	0.837
9	Cross	0.7632	0.703
10	Deaf	0.8382	0.876
11	Explore	0.7723	0.645
12	Hunt	0.8416	0.868
13	Knife	0.8716	0.868
14	Relay	0.8156	0.864
15	Reverse	0.8644	0.894
16	Role	0.8564	0.841
17	Tell	0.8302	0.879
18	Sad	0.8583	0.877
19	Student	0.8565	0.865
20	Table	0.7966	0.811

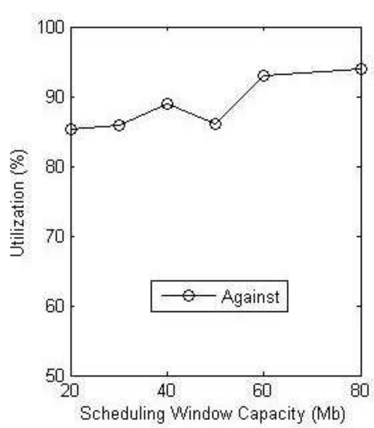

Fig. 17: Resource Utilization-Against

Table 4: Objective and Subjective Assessment Scores of Reconstructed Video

S.No	Videos	Objective Assessment (Recognition Rate)	Subjective Assessment (Likert Score)
1	Against	0.7321	0.782
2	Assume	0.7211	0.773
3	Aim	0.7215	0.821
4	Balloon	0.7752	0.831
5	Cake	0.7364	0.795
6	Chair	0.7661	0.772
7	Computer	0.77	0.857
8	Concentrate	0.7798	0.724
9	Cross	0.6987	0.603
10	Deaf	0.7281	0.745
11	Explore	0.6713	0.594
12	Hunt	0.7615	0.734
13	Knife	0.7716	0.768
14	Relay	0.7165	0.724
15	Reverse	0.7613	0.724
16	Role	0.7864	0.765
17	Tell	0.7325	0.779
18	Sad	0.7545	0.766
19	Student	0.7454	0.725
20	Table	0.6966	0.713

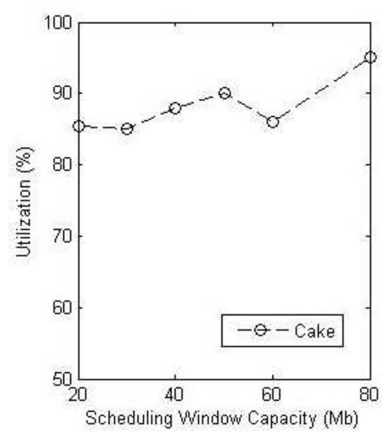

Fig. 18: Resource Utilization-Cake

6.2.3 Resource Utilization:

The resource utilization of the proposed heuristics is shown from Fig. 17 to Fig. 21 for different scheduling window capacity sizes. The capacity utilization is computed as follows: (specified in equation 20) If $r_0, r_1, \ldots, r_k$ are the data rates of the chosen layers, then

$$\text{Capacity Utilization} = \frac{\text{Total data sent in the schedule}}{\text{Total schedulable data capacity of a scheduling window}} \tag{20}$$

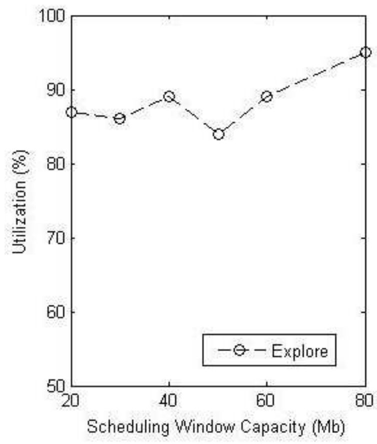

Fig. 19: Resource Utilization-Explore

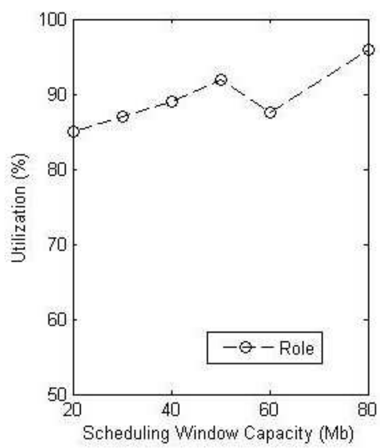

Fig. 20: Resource Utilization-Role

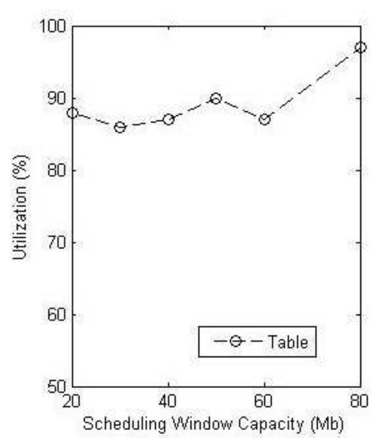

Fig. 21: Resource Utilization-Table

The total data sent in the schedule is calculated as $\sum \tau\, Dr_k$ and the total schedulable data capacity of a scheduling window is calculated as $D * F$ where F is the number of MBS areas, D is the data capacity of the MBS areas and τ denotes the duration of each TDD frame.

From the above observations, we therefore conclude that our heuristics can support large scale wireless streaming services and hence is more suitable for gesture based video conferencing applications.

7 Conclusion:

A framework using a Bandwidth Aware Gesture Based Layered (BAGBL) approach is proposed for catering the bandwidth requirements of video conferencing applications. We used an adaptive algorithm for extracting the key frames from the gesture video based on share energy trajectory to provide adaptive content delivery in time varied networks. We mathematically analyzed the problem of selecting the optimal sub layers of the layered video streams under bandwidth constraints and proposed a pair of new heuristics viz., MPMHSS and MDMP, one for substream scaling and the other for sub stream scheduling. Improving video quality at the subscriber device results in increased user satisfaction and in turn increases the recognition rate. Our solution can be used to transmit a larger number of videos at acceptable levels of quality. This also increases user satisfaction by offering the user more options in terms of content. BAGBL provides a versatile and flexible framework that can be tuned to the problem at hand i.e., this framework can be applied for other multimodal based real time applications as well. When tuned towards analysis efficiency, BAGBL can find good quality solutions in real time, within a predefined time bound.

Acknowledgement

Part of this work is supported by grants ISRO/RES/3/635/12-13 from the Indian Space Research Organization (ISRO), Ahmadabad, India.

References

[1] K. Finn, A.J. Sellen, S.B. Wilbur, Video-Mediated Communication. Lawrence Erlbaum, 323-347 (1997).
[2] Cisco WebEx Network Bandwidth, White Paper, (2013).
[3] H. Schwarz, D. Marpe, T. Weigand, Overview of the Scalable video coding extension of the H.264/AVC standard, IEEE Trans. Circuits Syst. Video Technol. **17**, 1103-1120 (2007).
[4] Nidal Nasser, Lutful Karim, and Tarik Taleb, Dynamic Multilevel Priority Packet Scheduling Scheme for Wireless Sensor Network, IEEE Transactions on Wireless Communications, **12**, 1448-1459 (2013).

[5] H. Cooper, R. Bowden, Large lexicon detection of sign language. Human Comput. Interact., 8897 (2007).
[6] F. Lichtenauer, E.A. Hendriks, M.J.T. Reinders, Sign language recognition by combining Statistical DTW and independent classification, IEEE Trans. Pattern Analysis and Machine Intelligence 10.1109/TPAMI.2008.123J. **30**, 2040-2046 (2008).
[7] Zhiyuan Lu, Xiang Chen, Qiang Li, Xu Zhang, Ping Zhou, A Hand Gesture Recognition Framework and Wearable Gesture-Based Interaction Prototype for Mobile Devices, IEEE Transactions on Human-Machine Systems, **44**, 293-299 (2014).
[8] H. Brashear, T. Starner, P. Lukowicz, H. Junker, Using multiple sensors for mobile sign language recognition, Proc. IEEE 7th Int. Symp. Wearable Comput., 4552 (2005).
[9] G. Fang, W. Gao, D. Zhao, Large-vocabulary continuous sign language recognition based on transition-movement models, IEEE Trans. Syst., Man, Cybern. A Syst. Humans, **37**, 19 (2007).
[10] S. Mitra, T. Acharya, Gesture recognition: A survey, IEEE Trans. Syst., Man, Cybern. C Appl. Rev. **37**, 311-323 (2007).
[11] V.E. Kosmidou, L.J. Hadjileontiadis, S.M. Panas, Evaluation of surface EMG features for the recognition of American sign language gestures. Proc. IEEE 28th Annu. Int. Conf. EMBS, 6197-6200 (2006).
[12] V.E. Kosmidou, L.J. Hadjileontiadis, Sign language recognition using intrinsic mode sample entropy on sEMG and accelerometer data. IEEE Trans. Biomed. Eng., **56**, 2879-2890 (2009).
[13] Yun Li, China Xiang Chen, Xu Zhang, Kongqiao Wang, Z.J. Wang, A Sign-Component-Based Framework for Chinese Sign Language Recognition Using Accelerometer and sEMG Data. IEEE Transactions on Biomedical Eng., **59**, 2695-2704 (2012).
[14] V.E. Kosmidou, L.J. Hadjileontiadis, Using sample entropy for automated sign language recognition on sEMG and accelerometer data. Med. Biol. Eng. Comput., **48**, 255-267 (2010).
[15] I. SushmitaMitra, Tinku Acharya, Gesture Recognition: A Survey. IEEE Transactions on systems, man, and cyberneticspart c: applications and reviews, **37**, 311-324 (2007).
[16] X. Chai, et al. Sign Language Recognition and Translation with Kinect. (2013).
[17] Li-Chun Wang, Ru Wang, De-Hui Kong,Bao-Cai Yin, Similarity Assessment Model for Chinese Sign Language Videos. IEEE Transactions on Multimedia, **16**, 751-761 (2014).
[18] J. Tran, Towards Human Centered Optimization of Mobile Sign Language Communication, Accessibility and Computing: The Newsletter of ACM SIGACCESS, **105**, 29-33 (2013).
[19] Bogdan Ciubotaru, Gabriel-MiroMuntean, GheorghitaGhinea, Objective Assessment of Region of Interest-Aware Adaptive Multimedia Streaming Quality, IEEE Transactions on Broadcasting, **55**, 202-212 (2009).
[20] J. Padhye, J. Kurose, D. Towsley, R. Koodli, A model based TCP friendly rate control protocol, Proc. NOSSDAV, (1999).
[21] D. Sisalem, A. Wolisz, LDA+ TCP-friendly adaptation: A measurement and comparison study. Proc. NOSSDAV, (2000).
[22] R. Rejaie, M. Handley, D. Estrin, RAP: An end-to-end rate-based congestion control mechanism for realtime streams in the Internet, Proc. INFOCOM, New York, USA, **3**, 1337-1345 (1999).
[23] L. Qiong, M. Van der Schaar, Providing adaptive QoS to layered video over wireless local area networks through real-time retry limit adaptation, IEEE Trans. Multimedia., **6**, 278-290 (2004).
[24] S.H. Shah, K. Chen, K. Nahrstedt, Dynamic Bandwidth management for single-hop Ad Hoc wireless networks, Proc. IEEE Intl. Conf. Pervasive Comp. and Commun. 195-203 (2003).
[25] K.A. Cha Content complexity adaptation for MPEG-4 audio-visual scene. IEEE Trans. Cons. Elec., **50**, 760-765 (2004).
[26] M. Van der Schaar, H. Radha, Adaptive motion-compensation finegranular-scalability (AMC-FGS) for wireless video, IEEE Trans. Circuits Syst. Video Technol., **12**, 360-371 (2002).
[27] N. Cranley, L. Murphy, P. Perry, Optimum adaptation trajectories for streamed multimedia, ACM Multimedia Systems Journal, **10**, 392-401 (2005).
[28] IEEE Standards for Local and Metropolitan Area Networks: Specific requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications Amendment 8: MAC Quality of Service Enhancements, IEEE Std 802.11e, (2005).
[29] G. Yang, T. Sun, M. Gerla, M.Y. Sanadidi, and L.J. Chen, Smooth and efficient real-time video transport in the presence of wireless errors, ACM Trans. Multimedia Comput., Commun.,Applicat., **2**, 109-126 (2006).
[30] J. Chakareski, P. Frossard, Rate-distortion optimized distributed packet scheduling of multiple video streams over shared communication resources, IEEE Trans. Multimedia. **8**, 207-218 (2006).
[31] L. Zhou, X. Wang, W. Tu, G. Muntean, B. Geller, Distributed scheduling scheme for video streaming over multi-channel multi-radiomulti-hop wireless networks, IEEE J. Select. Areas Commun. **28**, 409419 (2010).
[32] X. Zhu, R. Pan, N. Dukkipati, V. Subramanian, F. Bonomi, Layered internet video engineering (LIVE): Network-assisted bandwidth sharing and transient loss protection for scalable video streaming, Proc. IEEE Int. Conf. Computer Communications Mini-Conference, (2010).
[33] X. Zhu, B. Girod, Distributed media-aware rate allocation forwireless video streaming, IEEE Trans. Circuits Syst. Video Technol., **20**, 14621474 (2010).
[34] S. Khan, S. Duhovnikov, E. Steinbach, W. Kellerer, MOS-based multiuser multiapplication cross-layer optimization for mobile multimedia communication, Adv. Multimedia. 111 (2007).
[35] A. Ksentini, M. Naimi, A. Gueroui, Toward an improvement of H.264 video transmission over IEEE 802.11e through a cross-layer architecture, IEEE Commun. Mag., **44**, 107-104 (2006).
[36] M. van der Schaar, N. Shankar, Cross-layer wireless multimedia transmission: Challenges, principles and new paradigms.IEEE Wireless Commun., **12**, 50-58 (2005).

[37] H. Zhang, Y. Zheng, M. Khojastepour, S. Rangarajan, Cross-layer optimization for streaming scalable video over fading wireless networks, IEEE J. Select. Areas Commun. **28**, 344353 (2010).
[38] G. M. Muntean, P. Perry, L. Murphy, A new adaptive multimedia streaming system for All-IP multi-service networks, IEEE Trans. Broadcasting. **50**, 110 (2004).
[39] J. Chakareski, In-network packet scheduling and rate allocation: Acontent delivery perspective, IEEE Trans. Multimedia, **13**, 1092-1102 (2011).
[40] Hao Hu, Xiaoqing Zhu, Yao Wang, Fellow, IEEE, Rong Pan, Jiang Zhu, FlavioBonomi, Proxy-Based Multi-Stream Scalable Video Adaptation over Wireless Networks using Subjective Quality and Rate Models, IEEE Trans. Multimedia, **15**, 1638-1652 (2013).
[41] B.F. Momin, S.B. Pawar, Key Frame Extraction Using Features Aggregation, International Journal of Recent Development in Engineering and Technology, **2**, 81-85 (2014).
[42] M. Jones, and J. Rehg, Statistical color models with applications to skin detection, International Journal of Computer Vision, **46**, 81-96 (2002).
[43] Nikolaos Doulamis, Anastasios Doulamis and Dimitrios Kosmopoulos, Content-based Decomposition of Gesture Videos, IEEE Workshop on Signal Processing Systems, Design and Implementation, 319-324 (2005).
[44] M. Geilen, T. Basten, B.D. Theelen, and R. Otten, An algebra of Pareto Points, FundamentaInformaticae. **78**, 35-74 (2007).
[45] Y. Toyoda, A simplified algorithm for obtaining approximate solution to zero-one programming problems, Management Science, **21**, 1417-1427 (1975).
[46] Anbin Kim, Chanhong Park, Seong-Ho Jeong, Performance evaluation of downlink scheduling algorithms for video/voice transport over wireless networks, Ubiquitous and Future Networks (ICUFN), Fifth International Conference on, 791-794 (2013).
[47] F. Capozzi, G. Piro, L.A. Grieco, G. Boggia, P. Camarda, Downlink Packet Scheduling in LTE Cellular Networks: Key Design Issues and a Survey. Communications Surveys & Tutorials, IEEE., **15**, 678-700 (2013).
[48] A.A. Melkman, On-line construction of the convex hull of a simple polyline, Information Processing Letters, **25**, 11-12 (1987).
[49] SomsubhraSharangi, Ramesh Krishnamurti, Mohamed Hefeeda, Energy efficient Multicasting of Scalable Streams over WiMAX networks, International journal of Multimedia transmission, **13**, 102-115 (2011).
[50] Q. Munib, M. Habeeb, B. Takruri, AI-Malik, American sign language (ASL) recognition based on Hough transform and neural networks, Expert Systems with Applications, **32**, 24-37 (2007).
[51] http://users.iit.demokritos.gr/∼dkosmo/downloads/gesture /help_gestureDB.htm
[52] Y. Wang, Survey of Objective Video Quality Measurements, Tech report, Worcester Polytechnic Institute (2006).
[53] Definition of Quality of Experience (QoE), Reference: TD 109rev2 (PLEN/12), ITU International Telecommunication Union.

Swarna Parvathi is an Associate Professor in the department of Information Technology, Sri Venkateswara College of Engineering, Chennai. She received her B.E degree in Electronics and Communication Engineering from Manonmanium Sundaranar University, Tirunelveli in 1998 and her M.E in Communication Systems from Anna University, Chennai in 2004. At present, she is pursuing her Ph D in Computer Science and Engineering at Anna University, Chennai. Her research interests include Multimedia networks, Mobile Computing, network emulation, design and analysis.

K. S. Easwarakumar is a Professor at the Department of Computer Science and Engineering at Anna University, Chennai. He received his M.Tech in Computer and Information Sciences from Cochin University of Science and Technology, Cochin and Ph.D in Computer Science and Engineering from Indian Institute of Technology, Madras. His research interests include parallel and distributed computing, Data Structures and Algorithms, Graph Algorithms, Parallel Algorithms, Computational Geometry, Theoretical Computer Science and Molecular computing.

N. Devi is an Assistant Professor in the department of Information Technology, Sri Venkateswara College of Engineering, Chennai. She received her B.E degree in Computer Science and Engineering from Annamalai University, Chidhambaram in 1998 and her M.E in Multimedia Technology from Anna University, Chennai in 2003. At present, she is pursuing her Ph D in Computer Science and Engineering at Anna University, Chennai. Her research interests include Multimodal interaction and Image Processing.

Raju Das is a Senior Scientist in the Space Applications Center, Indian Space Research Organization (ISRO), Ahmadabad, India. He received his B.E. degree in Computer Science and Engineering from the National Institute of Technology, Assam, India in 1998. His research interests include Protocol design for Satellite Communication Network, Efficient & Reliable Data Communication, Mobility Management for Next Generation Mobile Satellite Networks, Dynamic Bandwidth Management, Earth Station Automation, Emergency Communication Network for Disaster Management and various Satellite based Applications developments like Tele education, Portable/Handheld multimedia terminal, Network Management System etc.