Dhaka Univ. J. Eng. & Tech. Vol. 1(2) 147-150, 2010 (January)

Video Conferencing over Very Narrow Band Internet Using Image Metamorphosis

A. S. Md. Mukarram Hossain, Shaily Kabir and Md. Haider Ali

Department of Computer Science and Engineering, University of Dhaka, Dhaka, Bangladesh

mukarram@cse.univdhaka.edu, shailykabir@yahoo.com, haider@univdhaka.edu

Received on 13.03.2010. Accepted for publication on 02.02.2011

Abstract

A *Video Conference* is a set of interactive telecommunication technologies which allows two or more locations to interact via two-way video and audio transmission simultaneously. *Image Metamorphosis* or image morphing is a powerful tool for creating visual effects to produce a fluid like animation between the images where the first image is viewed as transforming into the second image. In this paper, an efficient algorithm for video conferencing over very narrow band Internet is proposed using image morphing to reduce the amount of data transmission and the bandwidth requirement. The proposed system captures and extracts facial features of the human face and sends them to the receiver. Morphing algorithm at the receiver uses this information to generate the visual animations for display. The potency of the proposed algorithm has proved by the experimental results which show the reduction of the data transfer rate as well as bandwidth requirements during a two way video conferencing process. Experimental results also show that, the proposed idea can also be extended and implemented for compression and decompression of non-real time video applications like TV news, talk shows, in virtual class rooms and in distant learning.

Keywords : Video conference, image morphing, expression animation, video compression.

1 Introduction

Video conferencing uses telecommunications of audio and video to bring people at different sites together for a meeting. This can be a conversation between two people (point-to-point) or can involve several sites (multi-point). Besides the audio and visual transmission of meeting activities, video conferencing can be used to share documents, computer-displayed information and white boards.

Inherently video conferencing requires communication links with high data transmission capability. The usual frame rate for a video conference system is 25 Frames/s. Each frame contains bulk video data and for smooth conferencing, these data needs to be present at the other site. So, the bandwidth requirements are pretty high for such real time multimedia applications. The trend of using Internet to share multimedia is increasing rapidly and this eventually is reducing the share of bandwidth for all. In countries like Bangladesh, people badly suffer from inadequate bandwidth and are often unable to experience smooth video conferencing. The idea of distant learning and virtual class rooms are hence not a reality. So we need to devise systems that can achieve transmission of multimedia content over low bandwidth networks without loosing the quality and thus user satisfaction.

In video conferencing, the frontal face of the user is normally the only moving object of the video sequence. People notice the changes of the frequently changing object (face in this case) during a video conversation. So, the changes in the facial region are enough to describe the frame accurately. Human perception of face is generally based on some landmarks or facial features such as eye, nose, mouth, eyebrows etc. Small changes in feature regions have significant impact on the look of the person. So it is very important to be able to track these facial features. Eye corner points, nose tip, lip corner points, eyebrow boundaries are the most important features of a face. These points are referred to as *control points* throughout the text.

An efficient algorithm for smooth video conferencing over narrow band Internet is presented in this paper. The algorithm uses control point detection of the facial features [1] to determine the information required to send to the other site. Image morphing [2] then uses this information to animate the frames on the other site. Up to this point, the audio transmission has not been discussed. The amount of audio data that a multimedia system or an audio-visual system produce is minimal compared to the video data. Normally audio and video data are transmitted simultaneously with separate channel. Audio compressions like MPEG / audio compression [3] is commonly used in multimedia applications. A detail discussion of different audio compression techniques can be found in [4] and other literatures.

2. Related Works

Video Conference and Image Morphing technologies both employ complex and diversified mechanisms for their working. Many standards have recently been developed for video conferencing. The International Telecommunications Union (ITU) is the organization that creates these standards. The first standard developed, and the standard that most other standards are an extension of, is called H.320. This standard includes specifications for audio, video, and data transmission.

Present video conferencing systems mostly use compression to reduce the bandwidth requirement. Some of the popular and widely used compression techniques for still images include JPEG (Joint Photographic Experts Group) [5] and GIF (Graphics Interchange Format). For Video compression, popular compression techniques include Motion-JPEG [6] which is a *de facto* standard; MPEG-1 can encode video streams of 1.5Mbps, which is sufficient for CD-ROM applications. MPEG-1 has successors like MPEG-2, MPEG-4, and H.261. H.261 has been followed by

H.263 and H.264 standard. A detail discussion and comparison between various image and motion compression techniques is presented in [7].

Image metamorphosis or image morphing is a powerful tool for visual effects. The earliest attempts of using image metamorphosis in video include Michael Jackson's famous MTV music video **Black or White** which gave people some astonishment. Morphing is achieved by coupling image warping with color interpolation. The paper [2] describes the advancement in image morphing in term of addressing feature specification, warp generation and transition control. Among the most widely used morphing techniques, two are based on mesh warping [8] and field morphing [9].

Many commercial video conferencing systems have been developed by companies. One such commercial video conferencing system has been developed by PolycomTM (www.polycom.com). This system is specifically designed for broadband networks and thus cannot be used with narrow band Internet. Two researchers from University of Washington and Microsoft Corporation have proposed a video conferencing system using image morphing for low bandwidth network [10]. They proposed a face video transmission strategy, which can, in real-time, compress and transmit face video at very low bit-rates, such as 8Kb/s. They also presented the novel compression strategy: instead of compressing and transmitting every frame at a very low quality, they reduce the video frame-rate and send only important regions of carefully selected frames.

3. Methodologies

Current researchers have started thinking of using image morphing as a means for video compression. The studies suggest that image morphing can be an excellent choice for video compression over narrow band Internet. On view of that, a novel approach of bandwidth efficient video conferencing system is proposed here. This proposed system uses image morphing as a powerful tool for both video compression and animation. It is comprised of two processes running simultaneously in each of the participant's (sender and receiver) computer.

Sender Process

The first job of the sender process is to capture video frames regularly. Any standard video capturing device like digital camera, video camera or webcam can be used for this purpose. After capturing frames, the relative deviation between the current frame and the previous frame is computed using Equation 1.

$$D_r = \sum_{i=0}^{n-1}\sum_{j=0}^{h-1}\left[\left(R_{i,j(R)} - C_{i,j(R)}\right) + \left(R_{i,j(G)} - C_{i,j(G)}\right) + \left(R_{i,j(B)} - C_{i,j(B)}\right)\right] \quad \ldots\ldots\ldots(1)$$

The relative frame deviation value, D_r, gives an estimation of how much changes have occurred after the last transmission of frame data. The pixel wise deviations are stored in a buffer named *pixelDifference* to be used later. After calculating the deviation, the value is compared with a threshold T_H. If the value exceeds the threshold ($D_r > T_H$), then complete frame is sent to the receiver process. Before sending the full frame information, the control points of the frame are calculated using the method presented by [1]. If the relative deviation does not exceed T_H, then there can be two possibilities. First, the frame has changed very little, if any and thus the relative deviation is below another threshold T_L which means $0 < D_r < T_L$. This can be a result of slight change of the position of the person facing the camera, very little movement of eyes or lips, change of the room light etc. The experimentation shows that, this is the type of frame that is encountered most frequently. We do not need to transmit these frames as they contain information that are of little or no use. The sender simply drops the frame and adds 5 to the count *frameCount* that keeps track of the number of frames not transmitted since last. After dropping the frame, sender process captures the 5th frame again and the whole process is continued.

There are some frames with relative deviation $T_L < D_r < T_H$. With relative difference value between T_H and T_L, the frame is not sent full or dropped at all. The relative deviation value indicates that the frame contains changes that are important factors to consider. For this type of frames, we consider the buffer *pixelDifference* that is holding the current differences of RGB values for every pixel. Then the Process calculates the control point for the current frame. This control information is then sent with the contents of *pixelDifference*. We use a heuristic here, rather than sending complete buffer, we send pixel values having difference greater than a certain limit. The experimentation shows that, we can easily ignore frames having change within 20-30 pixel level. With heuristics applied, the number of pixels to send reduces dramatically. This eventually reduces the amount of information transferred between communicating parties reducing the bandwidth requirements.

Receiver Process

The receiver process module is fairly simple compared to sender process module. Its main duty is to capture data sent by the sender process and generate intermediary frames for animation. Sender sends two kinds of frame information:

- Full frame information along with feature correspondences.
- Control points for frames only.

The receiver process has been designed to handle transmission of both kind of frame information. This process is always waiting to receive frame information from the sender. In both the cases, *frameCount* is sent that holds the number of frames dropped. This value is used to animate correct number of frames from the reference frame and the received information. After receiving the frame information and the control points, morphing is performed to animate the video sequence.

Many image morphing techniques are available now for instance, like Mesh Warping [8], Field Morphing [9], Free-form Deformation, View Morphing [11] etc. Recently [12] showed the Relationship between Full Width Half Maximum (FWHM) and Human Facial Shape Distortion in

Image Metamorphosis process. The receiver process uses the morphing technique based on the metamorphosis method. After receiving each frame, the receiver process will have two frames in hand: the source frame F_s and the target frame F_t. Morphing process is done using forward mapping where F_s is gradually distorted and turned into F_t. The positions and the colors of pixels in the two frames are interpolated for producing transitions between themselves. The pixel correspondence from F_s is denoted P_s and F_t is denoted P_t. For each pixel coordinate, the displacement between the source and target is calculated using Equations 2 and Equations 3 until P_s reaches P_t.

$$M_X = P_{t(x)} - P_{s(x)} \quad \text{.....(2)}$$

$$M_Y = P_{t(y)} - P_{s(y)} \quad \text{.....(3)}$$

After calculating the displacement value for each pair of pixels, the distance between the current pixel (X,Y) and $(P_{s(x)},P_{s(y)})$ is calculated using Equation 4.

$$pixelDis\tan ce = \sqrt{\left(P_{s(x)} - X\right)^2 + \left(P_{s(y)} - Y\right)^2} \quad \text{....(4)}$$

The Gaussian distribution pattern is then applied to the unconstrained pixel positions to calculate the pixel movements. Pixels closer to the predefined source pixel are affected most and more distant pixels are affected least. Movement of the pixels is calculated using Equation 5. In the given equation, FWHM is the fall off value that determines the effect of distortion for a pair of pixel coordinate. The value of FWHM ranges between 10 and 200, but [12] suggested the FWHM to be 90 for better morphing.

$$Movement(X,Y) = \left(M_X, M_Y\right) \times e^{-\left(\frac{pixelDis\tan ce}{FWHM}\right)} \quad \text{....(5)}$$

4. Experimental Results

Several experiments have been carried out with the proposed system. For each session of video conferencing, the frames were captured and stored as AVI (Audio Video Interleave) file. Then the experimentations were performed on those files. The experimentations include: 1) Detecting the accuracy of the control point detection algorithm, 2) Calculating the average color deviation between frames, 3) Monitoring the relative displacement of feature points, 4) Estimating the amount of data required for frame with: little, moderate, and with extreme deviation, 5) Evaluating the performance of morphing, 6) Estimating the bandwidth requirement of the video conferencing system.

The control point detection (feature extraction) system has been tested on many still images and video frames. The still images are all in BMP format and the video frames are captured and stored as BMP images. Figure 1 shows an accurate detection of control points on the left image. The right image shows another frame of the same person where the facial feature points is partially detected. False detection is due to the orientation of the face and the background.

Fig. 1: Accurate detection of the facial features (control points)

The experiments have been performed on captured video files that were stored as AVI files. After saving the video file, the average color deviations of the subsequent frames were calculated and plotted on charts to view the relative changes. These changes give an estimation of the nature of the frame sequences. Small deviation refers to the subtle changes on the subsequent frame. The results have been plotted on a graph shown in Figure 2 to demonstrate the average color deviation between frames of the video file *video1.avi*.

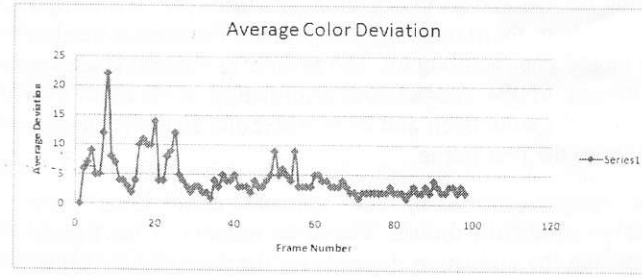

Fig. 2: Average color deviation between subsequent frames in *video1.avi*

The displacement of the facial features for subsequent frames have also been computed and plotted on a second graph. Figure 3 displays the displacement of facial feature points for the video file *video.avi*. This graph shows the changes of the facial feature points where sharp edges represent the abrupt change in the person's facial features.

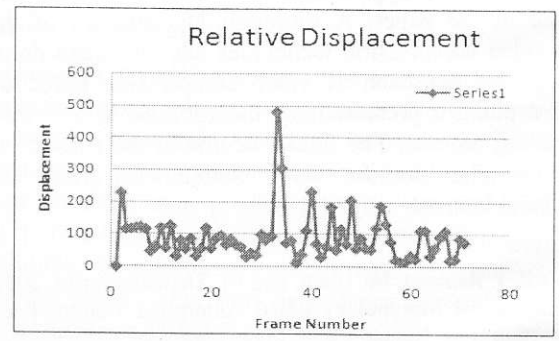

Fig. 3: Graph showing relative displacement between frames in *video.avi*

Generating Facial expressions are the most challenging task of any video animation. In video conferencing, the facial features are extracted to track the human facial expressions. These expressions include—eye opening and closing, mouth opening and closing during conversation, movement of the eyebrows, expressions showing emotions etc. The proposed algorithm is capable of tracking simple expressions and reproducing them using image morphing. The sample

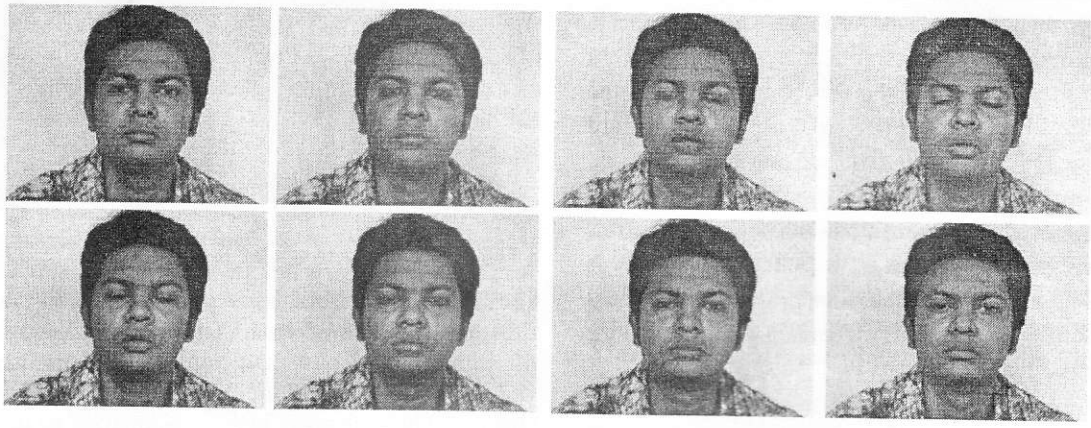

Fig. 4: Animated frame of the video sequence for mouth and eye opening, closing

animation given in Figure 4 is generated using the frame and feature specification supplied by the sender process. The animation starts from top-left frame where the person is having his eye and mouth open. Animations end at top right frame, where the frame has changed and the person has his eye closed and mouth open. The reverse animation is shown in images of the second row. Animation starts with eye closed and mouth open and ends where the frame is exactly similar to the first frame.

The morphing process can animate expressions with variable number of frames. The exact number of the frames producing the animation depends on the *frameCount* value. If *frameCount* holds value 5, then 5 frames will be used to animate the expression. The same stands for *frameCount* value of 10 and 15. So the speed of the animation can be controlled by the value of *frameCount*. It must be noted that, the sender process provides the value of the variable to the receiver process.

5. Conclusion

A new approach to video conferencing and similar real-time and non real-time applications has been presented and explained in the paper. A thorough investigation of the current video compression techniques has also been done. An alternative approach of video compression based on image morphing is presented and incorporated to the video conferencing process. The final outcome of the research is an efficient algorithm for video conferencing over very narrow band Internet.

References

1. M. Ali, I. Rahman, M. Islam, and M. Shahiduzzaman, 2007. Mathematical Morphology based Automated Control Point Detection from Human Facial Image, *Machine Graphics and Vision*, **16**(2), 153-162, 2007.
2. G.Wolberg, 1996. Recent Advances in Image Morphing, *In quot; Proc. Computer Graphics International.*
3. D. Pan, M. Inc, and I. Schaumburg, 1995. A Tutorial on MPEG/audio Compression, *IEEE multimedia*, **2**(2), 60-74.
4. D. Pan, 1993. Digital Audio Compression, *Digital Technical Journal*, **5**(2), 28-40.
5. G. Wallace *et al.*, 1991. The JPEG Still Picture Compression Standard, *Communications of the ACM*, **34**(4), 30-44.
6. T. Lane, *JPEG-faq, part 1.* ftp://rtfm.mit.edu/pub/usenet/news.answers/jpeg-faq/part1.
7. R. Aalmoes and P. Bosch, 1995. Overview of Still-picture and Video Compression Standards, *Pegasus Paper, Citeseer*, 95-98.
8. G. Wolberg, 1990. Digital Image Warping. IEEE Computer Society Press Los Alamitos, Calif.
9. T. Beier and S. Neely, 1992. Feature-based image metamorphosis, *SIGGRAPH Comput.Graph.*, **26**(2), 35-42.
10. J. Wang and M. Cohen, 2005. Very Low Frame-rate Video Streaming for Face-to-face Teleconference. *Data Compression Conference,Proceedings. DCC 2005.* 309-318.
11. S. Seitz and C. Dyer, 1996. View morphing. *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, 21-30.
12. S. Kabir and M. H. Ali, 2005. A Heuristic Approach of Establishing the Relationship between Full Width Half Maximum (FWHM) and Human Facial Shape Distortion in Image Metamorphosis. *Computer and Information Technology (ICCIT2005).*