Feeling the Motion of Object in a Dynamic Image Sequence through Haptic Interface

AnushaJayasiri, KatsuhitoAkahane, Makoto Sato

Precision and Intelligence Laboratory, Tokyo Institute of Technology
4259, Nagasstuta-cho, Yokohama-shi, 226-8503, Japan

{anusha, kakahane}@hi.pi.titech.ac.jp, msato@pi.titech.ac.jp

Abstract— **In this paperwe discussed how to incorporate haptic signals with a dynamic image sequence or other words a videoto feel the motion of objects in it.Haptic technologies are being used in a wide range of application areas. However, the incorporation of haptic interface technology into a video media is still in its infancy.With the invention of digital multimedia and immersive displays, significance of exploring new ways of interacting with video media has grown up. Rather than just seeing and hearing a video, viewers` experience can be further enhanced by letting them feel the movement of the objects in the video through haptic interface, as it is an additional sensation to seeing and hearing. The objective of this research is to use haptic interface technology to interact with a video and enable the viewers to feel the motion of objects in the video beyond passive watching and listening. In this paper, we discuss how to feel the motion, which is computed from frame to frame calculation of velocity using optical flow. For haptic motion rendering, we have proposed a method by experimentallyevaluating two methods using a gain controller and using a non-linear function. To interact with the video we used the string based haptic device, SPIDAR, which provides a high definition force feedback sensation to users.**

Keywords— **Optical Flow, Haptic Interface, Haptic Motion, SPIDAR (Space Interface Device for Artificial Reality)**

I. INTRODUCTION

Nowadays, there is a significant evolvement in haptic technologies and they are being used in different kind of application areas such as in training [3] [20], education [19], entertainment and 3D interaction [16]. Unlike traditional interfaces that provide visual and auditory information, haptic interfaces can generate mechanical signals, which stimulate human kinaesthetic and touch channels[11]. As a result it enables the users to touch and manipulate objects in the scene and can enhance human senses.

With the invention of digital multimedia and immersive displays, significance of exploring new ways of interacting with video media has grown up. For example, in three-dimensional television (3D-TV), viewers can see objects in true dimensions and natural color, providing a natural viewing experience with advanced audio technologies. The Viewers` experience can be further enhanced if viewers can get the feeling of the movement of the objects in a dynamic image sequence or in other word a video through haptic interface, as it is an additional sensation to seeing and hearing.

However, the incorporation of haptic interface technology into video media is still in its infancy. As Dinder et al [10] discusses, there are 3 types of haptic effects which are cooperating with haptic interaction with video media. Those are haptic structure, haptic texture and haptic motion. Haptic structure refers to the touching or getting the feeling of the geometry of an object in the video scene. Haptic texture refers to the rendering of surface properties such that roughness of various objects in the video scene. Haptic motion refers to the rendering of forces related to the moving objects in the scene. In this research we address the haptic motion part. Consequently, this research is an attempt to use haptic technology to interact with avideo with the objective of enabling the viewers to feel the motion force of objects in the video beyond simply passive watching and listening. Here we used 2D video representation, which includes only one moving object for rendering haptic motion.

To achieve this, we have proposed a method by evaluating two methods;one usinga gain controller and another using a non-linear functionby using a haptic device SPIDAR, which is developed by Sato laboratory, Tokyo Institute of Technology. SPIDAR, which stands for `SPace Interface Device for Artificial Reality`, is a string-based haptic device and can beused in various types of virtual reality applications ranging from desktop, workbench, human scale and networked environments. SPIDAR has three distinguishablefeatures: scalability, string based and transparency. Scalability means, SPIDAR can be fit to different working spaces such as desktop,workbench, or human-scalewithsimple modification on its structure layout.String basedtechnology gives the developer the ability todisplay position, orientation and force feedback at once andmaking it an effective mean for pointing and controlfor 3D environment. Transparency is based on its string-based technology, because SPIDAR keeps the working spacetransparent and do not obscure the visual display.Different versions of SPIDAR systems are available for simple `pick and place` task to more complicated physical based interactions [21].

For this research we used the SPIDAR-G system shown in Fig. 1, which is a grip type, tension based 6 degrees of freedom; 3 degrees of freedom for translation, 3 degrees of freedom for rotation and grasp enabled force-feedback device. This device has a grip and this grip is attached to 8 strings.

Each string is connected to a motor and an encoder at one end and tothe grip at the other end. The feedback force is determined by the tension of each string generated by the motor, which is transformedto the users hand through the grip. By connecting this device to a personal computer, it provides a high definition force feedback sensation to the user [1][14].
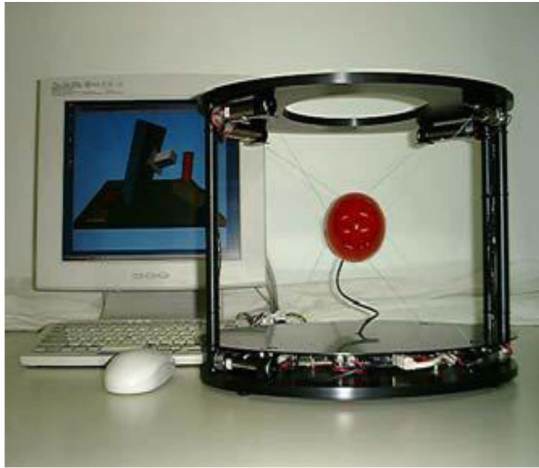

Fig. 1 SPIDAR-G haptic device

As a result, the viewercan feel the movement of the objects in the video or the forces related to the movement into his/her hand rather than just seeing it as illustrated in Fig. 2.

Fig.2Haptic interaction scenario

This paper discusseshow to associate haptic signals with video media to feel the motion of objects in it. The paper is organized as follows: section 2 presents related work in haptic rendering with video mediaand using SPIDAR for image haptization. Section 3 presents our proposed approach for feature point selection, feature point tracking, motion estimation and haptic motion rendering. We further elaborate on our proposed approach for haptic motion rendering using a gain controller and a non-linear function. Section 4 presents the experimental evaluation of the above two methodsto identify which method is better.Section 5 concludes the work and section 6 presents future work.

II. RELATED WORK

Incorporation of haptic technology into video media is not adequately researched in the haptic rendering field. This section summarizes few of them.

Dinder et al [10] have introduced the concept of haptic motion for the first time and they discussed a method to compute haptic structure and motion signals for 2D video –plus-depth representation which enables the viewer to navigate the scene and get the experience of the geometry of objects in the scene as well as the forces related to moving objects in the scene using PHANToM haptic interface. They model the total force as the sum of static and dynamic forces. While the static force is related to the geometry, material and surface properties of an object, dynamic force relates to the relative motion between the object and haptic interaction point.

O`Modhrainet al [17]have discussed how haptic interaction can enhance and enrich the viewer's experience in broadcast content. They proposed a touch TV project with the use of Gravis Xterminator Force and remote control handset to generate haptic cues for cartoons and live sports broadcastings which adds greater sense of immersion. They believe that the interactive nature of touch media has the potential to greatly enrich interactive TV by physically engaging the viewer in the programmed experience.

Cha et al [6][7][8] proposed a touchable 3D video system which provides haptic interaction with objects in a video scene. As a result it enables the viewers to actively touch a video scene through a PHANToMforce feedback device and enable to physically explore the video content and feel various haptic properties such that texture, height map and stiffness of the scene. They introduced Depth Image-Based Haptic Representation (DIBHR) method to add haptic surface properties of the video media.

Kim et al [13]proposed a 3DTV system which enables not only enjoying a high-quality 3D video in real time, but also user can experiencing various user-friendly interactions such as free viewpoint changing, composition of computer graphics and haptic display. They created a new representation of dynamic 3D scene, called 3D depth video and viewer can touch the shape of it by wearing a haptic device using a haptic rendering algorithm.

Because of the distinguishing features of scalability, string based and transparency made from SPIDAR system, it`s interface is used in various types of virtual reality systems [21]. Even though SPIDAR has not been used for haptization in video media,recently it has been used in the context of images. Liu et al [15] have proposed a haptization system which provides the users with sense of touch on an image with local deformations using SPIDAR haptic interface. In this research, user can interact with the image contents using SPIDAR-G haptic device.

III. PROPOSED APPROACH

The proposed approach for this research can be illustrated using a block diagram as in Fig. 3.
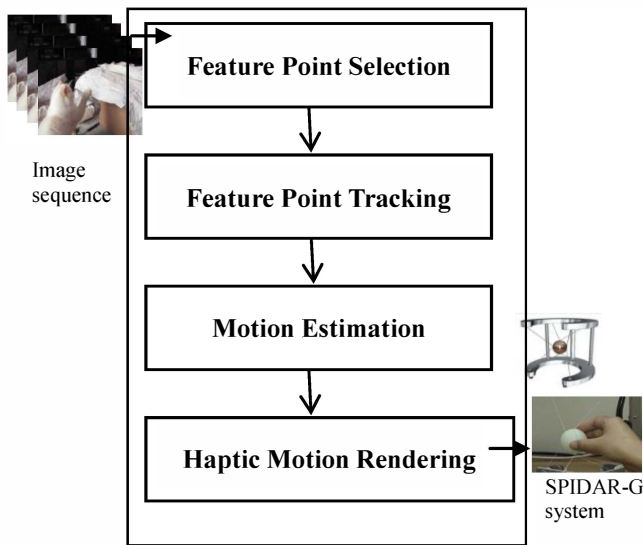

Fig. 3 Block diagram of the proposed approach

As shown in Fig. 3, it has four major parts: feature point selection, feature point tracking, motion estimation and haptic motion rendering. Videos are made up of individual images called frames. By using this concept, as the first thing it identifies good features of an image frame by frame and finally viewer can get the feeling of the objects in the video into his/her hand through the SPIDAR-G haptic interface. The following sections broadlydescribe the above parts.

A. Feature Point Selection

Feature points selection is an important task of any computer vision and image processing application. Since feature selection is the starting point of many computer vision

algorithms, the performance of the subsequent algorithm as well as the overall performance of the process basically depends on it.

Feature point selection finds which points are good to track. For example corners or good textures may be good feature points of an image. There are various methods existing for feature selection such as Harris, Canny, Sobel etc. Among those methods Shi & Thomasi algorithm gives better results than the others [2]. Besides, this algorithm is more efficient in interesting point detection and hence, the processing time of the overall process becomes less. Therefore we used Shi & Thomasi algorithm for feature points selection in the image sequence. This algorithm is based on the assumption that the brightness of a pixel does not change as it is tracked from frame to frame.

B. Feature PointTracking

Feature point tracking identifies above features reliably from frame to frame and those features are tracked using feature matching. Also we need to measure motion of the objects between two frames in an image sequence without any prior knowledge about the content of those frames. To address the above concerns, we used optical flow technique [4].

Optical Flow is the distribution of apparent velocities of movement of brightness patterns in an image. Optical flow arises from relative motion of objects and the viewer [12].The way an object moves when it is seen or followed in a video or sequence of images is known as optical flow [18]. There are two types of optical flow methods namely dense optical flow and sparse optical flow. In dense optical flow methods, it associates velocity with every pixel in an image. Horn-Schunck method and block matching method are examples for this type of optical flow [4]. In practice, calculating dense optical flow is not easy because of the high computational cost. Alternatively, sparse optical flow techniques calculate velocities only on the points which have certain desirable properties.
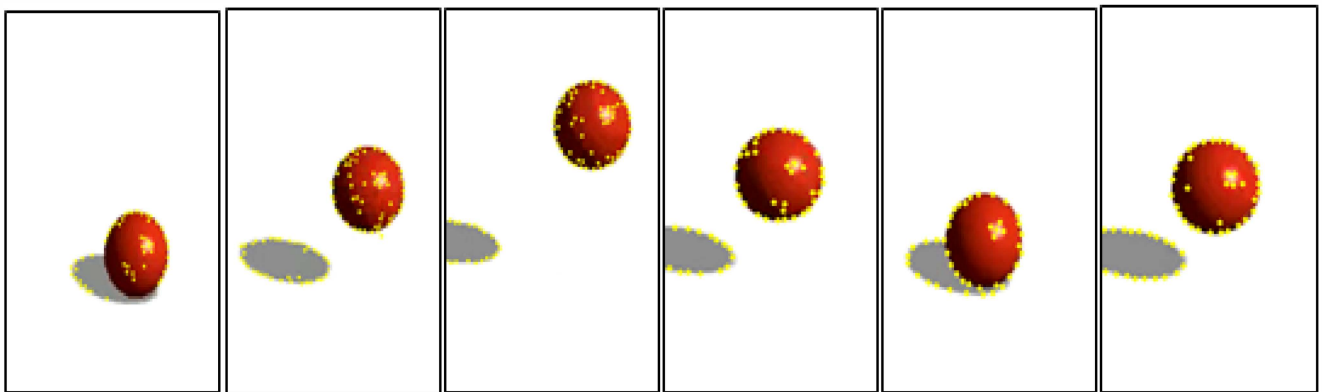

Fig. 4 Feature point selection with the Shi & Thomasi method and tracking of those points in the subsequent frames using the Pyramid Lucas-Kanade feature tracker
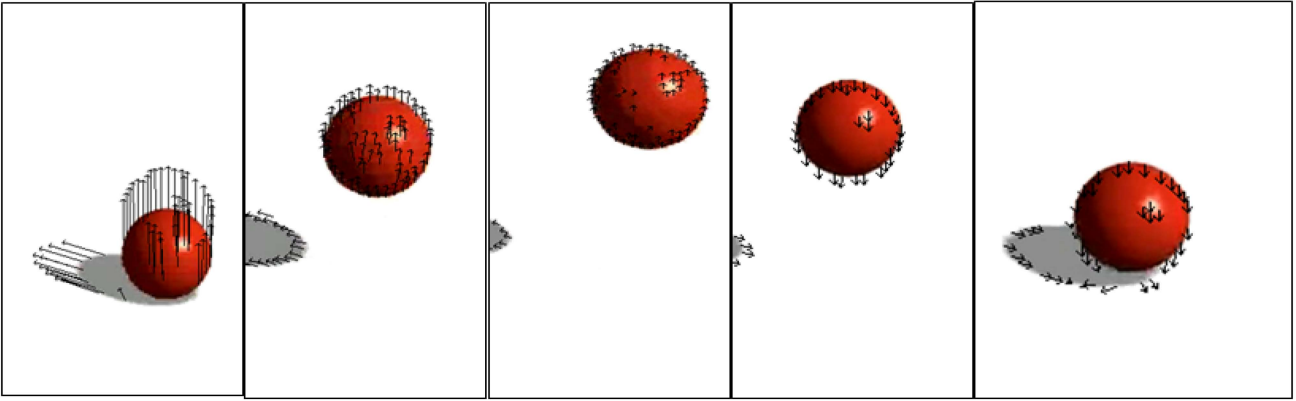
Fig. 5 Results of optical flow

Lucas-Kanade method is an example for this type of optical flow [4] [18].We used Pyramid Lucas-Kanade Algorithm, which is a pyramidal implementation of the Lucas-Kanade feature tracker [5]. At first in this technique, it solves optical flow at the top layer of the pyramid and then use the resulting motion estimates as the starting point for the next layer down. It continues going down the pyramid in this manner until it reaches the lowest level. Therefore by using this method, it can track faster and longer motions [4]. This has less computation and therefore it could be easily adapted for real time applications [2].

We used a synthetic scene of a bouncing ball video to evaluate our approach. Fig. 4 shows the obtained results for feature point selection and tracking with the use of Shi & Thomasi method and the Pyramid Lucas-Kanade method for the video sequence of the bouncing ball. Fig. 5 shows the optical flows of that image sequences. In that figure, the direction of each arrow represents the direction of optical flow and the length of each arrow represents the magnitude of the optical flow.

C. Motion Estimation

This section explains how we calculate the motion of an object in the image frame.

We use velocity of a feature point to estimate the motion. As shown in Fig. 6, the position of a feature point in two subsequent frames at time t and $t + \Delta t$ can be represented as $p_i(t)$ and $p_i(t + \Delta t)$.

The velocity of the feature point is calculated using the equation (1).

$$\overrightarrow{v_i(t)} = \frac{p_i(t+\Delta t)-p_i(t)}{\Delta t} \qquad (1)$$

Here, if there are N feature points in a frame, then the velocity of each frame is given by the average velocity of feature points in the image frame, as shown in equation (2).

$$\overrightarrow{v_f(t)} = \frac{1}{N}\sum_{i=1}^{N} v_i(t) \quad (2)$$
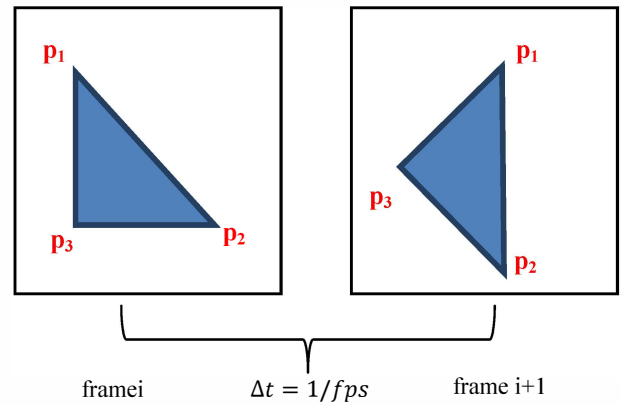

Fig. 6 Two consecutive frames of an image sequence

The Fig.7, 8 and 9 shows the changes of the position, velocity and acceleration of the ball in each frame during each run.

Bouncing ball is a good example for position, velocity and acceleration change against time because these graphs convey lots of information about speeding up, slowing down, rising or falling of the ball.

According to the Fig. 7, we can easily recognize the occasions when the ball is falling and rising. As shown in Fig. 8, when the ball is falling, the velocities on the graph are shown as negative values and the velocity is increasing. On the other hand, when the ball is rising, the velocities on the graph are shown as positive values and the velocity is decreasing. At the top of each bounce the velocity is zero because the ball changes its moving direction.

As the ball falls towards the floor its velocity increases and just before it hits the floor, its velocity becomes maximum.. Immediately after leaving the floor, i.e. at the start of the upward journey, its velocity is maximum and it is in the upward

direction. As the ball rises towards its highest position, its velocity approaches zero.

The Fig. 7, 8 and 9 complies with the physics of a bouncing ball [9] and hence we can conclude that our proposed method is accurate.
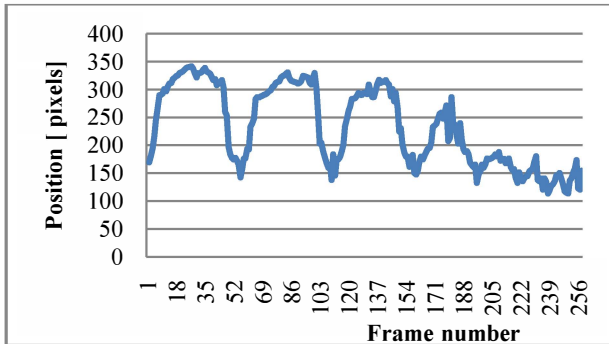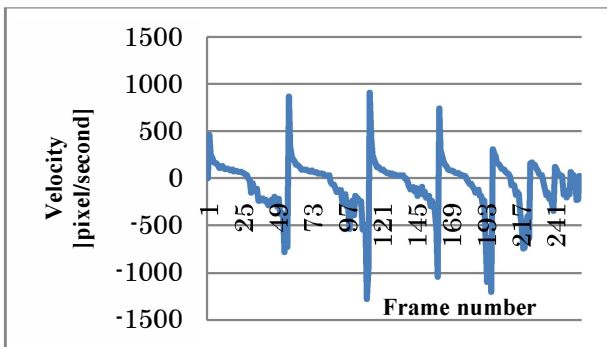

Fig. 7 Position changes in bouncing ball
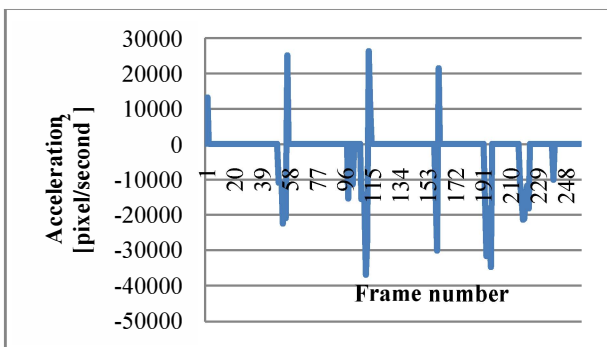

Fig. 8 Velocity changes in bouncing ball

Fig. 9 Acceleration changes in bouncing ball

D. Haptic Motion Rendering

Haptic motion rendering means rendering of forces related to the moving objects in the scene. In this section we explain how we calculated forces based on the above velocity changes in the video.

We used SPIDAR-G haptic device shown in figure 1. It generates 3 degrees of freedom force feedback by controlling the tension of each string in the system.

However, the high velocities produced by high force and low velocities produced by low force lead to an unrealistic sensation. To overcome this problem and to get a realistic sensation to the user we need to reduce force for high velocities and increase the force for low velocities. For this purpose, we evaluated two alternative methods using a gain controller and a nonlinear function to identify which one performs better.

I)   Method Using a Gain Controller:
Automatic gain controller is a feature found on many electric circuits that automatically controls the gain of a signal.

Using the gain controller method, the feedback force is calculated from equation (3). This enables user to get the feeling of the movement of the object.

$$\overrightarrow{F(t)} = k * \overrightarrow{v(t)} \qquad (3)$$

Here k is a gain controller.

Calculation of k is done as in equation (4) to control the feedback force within a sensible region for all velocity levels. In other words, the purpose of the gain controller k is to increase the feedback force for weak changes in velocity and decrease the feedback force for strong changes in velocity.

$$k = \frac{F_{max}}{V_{max}\ (T)} \qquad (4)$$

Here F_{max} is the maximum force output level of the SPIDAR-G for better sensation for this application. $V_{max}\ (T)$ is the maximum velocity of the dynamic image sequence at a time T, which can be expressed as in equation (5).

$$V_{max}\ (T) = \ \{\ V_{max}\ (t)0{\le}t{\le}T \quad (5)$$

We analysed the pattern of k value using the previously mentioned image sequence of the bouncing ball. Fig. 10 shows the results of the changing k value for the image sequence.
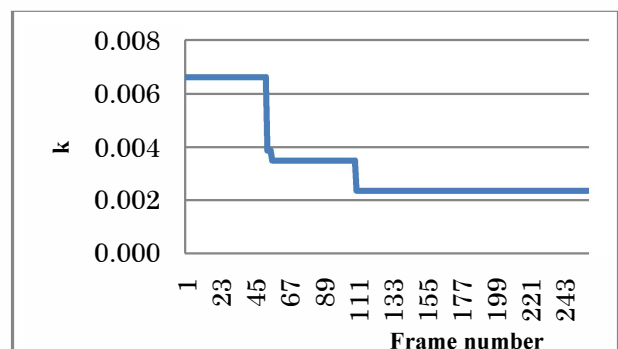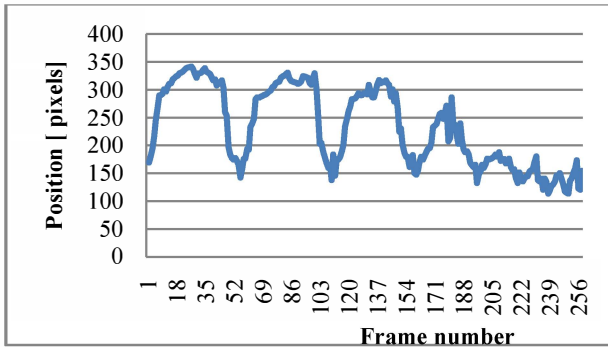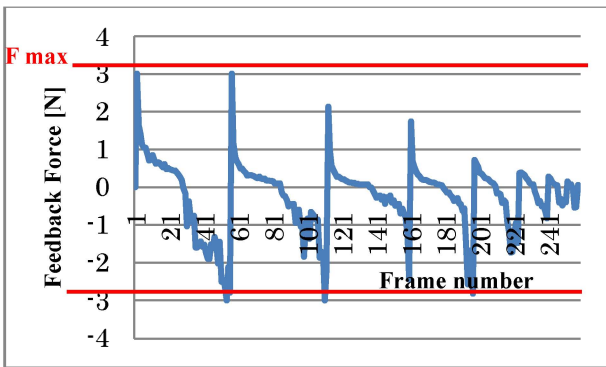

Fig. 10 Performance of the k value

Resulting force feedback of the SPIDAR for the image

sequence calculated using equation (3) is shown in Fig 11. Notably, resulting force is within the sensible region of the SPIDAR-G and hence user can smoothly feel the movement of the objects in the video.

(a)   Position changes of a bouncing ball

(b)Speed changes of a bouncing ball

Fig. 11 Feedback Force generated from SPIDAR forthe bouncing ball using the gain controller method

II) *Method Using a Nonlinear Function*

Similar to the previous method, the purpose of using a nonlinear function is to maintain the feedback force in the sensible region by decreasing the feedback force for high velocities and increasing the feedback force for low velocities. Fig. 12   illustrate this idea in a graphical form.
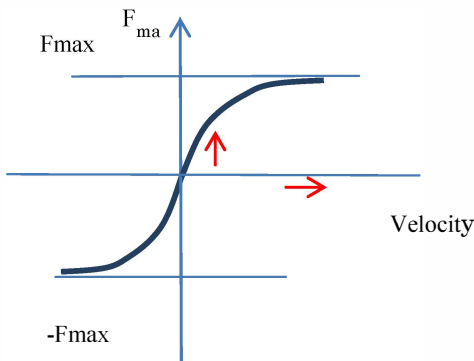

Fig.12 Purpose of the non-linear function

Using a non-linear function, the feedback force to sense the motion of objects is calculated from equation (6).

$$F(t) = f(v(t)) \quad (6)$$

To achieve the above objective, we tested with different sigmoid functions and selected the simple sigmoid function of inverse tangent function and the corresponding conversion is shown by equation(7).
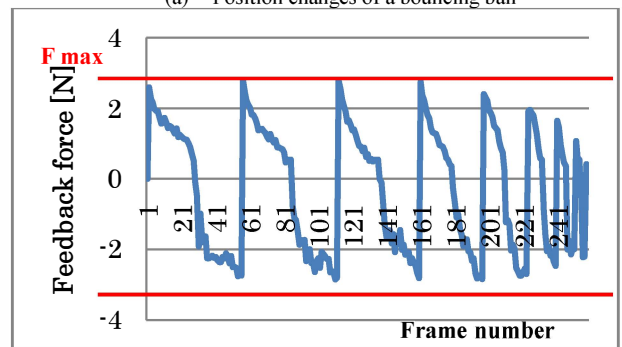
$$F(t) = \frac{2\,F_{max}}{\pi} \tan^{-1}(\alpha * V(t)) \qquad (7)$$

Here α is chosen as 0.01.

Resulting force feedback of the SPIDAR for the image sequence calculated using equation (7) is shown in Fig. 13. Notably, resulting force is within the sensible region of the SPIDAR-G and hence user can smoothly feel the movement of the objects in the video.
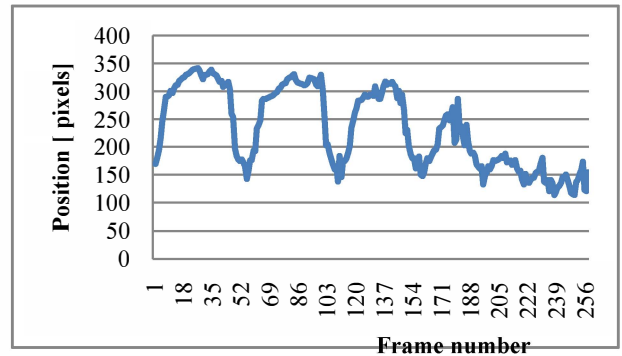
(a)   Position changes of a bouncing ball

(b)Speed changes of a bouncing ball

Fig. 13 Feedback Force generated from SPIDAR forthe bouncing ball using a non linear function

We can compare the resulting feedback force using two methods: a gain controller and non-linear function with respect to the position and velocity changes by using Fig. 11 and Fig. 13. It is clear that from Fig. 13, nonlinear function method outperforms the gain controller method,since it increases the feedback force for low velocities than the gain controller method and decreases the feedback force for high velocities

reasonably within the sensible region of the SPIDAR-G haptic device. In other words, user can feel the movement of the objects in the video even for smaller changes of the velocity. As a result user can get a continuous feeling of the movement of the ball even when theball is in the air.

IV. EXPERIMENTAL EVALUATION

In order to further evaluate which method gives better feeling about the movement of the objects in the video, we performed a qualitative experiment by getting the involvement of real users. For each user we conducted the experiment using twomethods, i.e gain controller method and the non-linear function method. We used a questionnaire to get the users feedback for each method. In the questionnaire the user was first asked to rate the two methods in a scale of `Bad`, `Poor`, `Average`, `Good` and `Very good` based on their feeling of movement of the object in the video.

For the analysis, we generally considered the responses of `Poor` and `Bad` as negative responses such that the feeling of the movement is not satisfactory for the respective method and denoted by `Bad`. On the other hand, the responses of `Average`, `Good` and `Very good` considered as positive responses such that the feeling of the movement is satisfactory for the respective method and denoted by `Good`.

The result of this rating is shown in Fig. 14. This graph shows how users rate each method based on their feeling as `Good ` or `Bad`. We tested with the involvement of ten numbers of real participants and it is clear thatmore than 60% of users responded that using a non-linear function is better than using a gain controller.

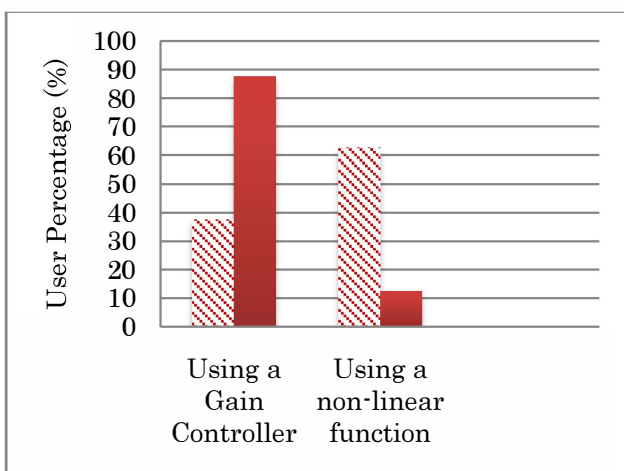|	Using a Gain Controller	Using a non-linear function
Good	37.5	62.5
Bad	87.5	12.5

Fig. 14 Result of responses of users for each method

Fig. 15 shows the distribution of positive responses for the non-linear function method, which was the selected method by most users as the better method out of the two.
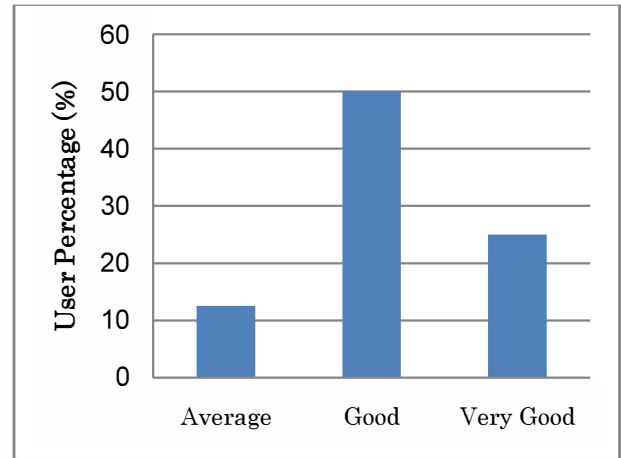

Fig. 15 Distribution of positive responses for using a non-linear function method

From Graph 15 it is clear that more than 70% of the users who selected this method as the better method have rated its performance as good or very good.

V. CONCLUSION

In thisresearch we concern how to associate haptic signals with a video to feel the haptic motion. To achieve the above purpose, we have experimentally evaluated two methods i.e.the gain controller method and the non-linear function method, with the objective of identifying the better one. We can conclude that using a nonlinear function is more effective than using a gain controller because user can get a continuous feedback force and hence, can get the continuous feeling of the movement of the objects in the video. As a result, user can feel the movement of the objects in the video as additional information than seeing it.

VI. FUTURE WORK

We tested our method with synthetic as well as real image sequences and identified that this is more effective for synthetic and real image sequence with less background noise.
We conclude this result by using one synthetic image sequence. In future we will evaluate the result by using movement of numerous types of synthetic and real image sequences.

Due to the higher background noise, this method is not suitable for real image sequences with object rich environment. In the future, we expect to improve and evaluate our method for real image sequences by eliminating the background noise.

Furthermore, since SPIDAR-G is a grip type haptic device, user can feel the movement of the object only to the

grippinghand. However this is not adequate to a realistic sensation. Therefore, another objective of our future work is to improve the system by identification of a suitable device such that it enables the user to get the whole body sensation.

REFERENCES

[1] K.Akahane,S. Hasegawa, Y. Koike, M. Sato, "A proposal of a High Definition Haptic Rendering for Stability and Fidelity", *Proceedings of the 16th International Conference on Artificial Reality and Telexistence Workshop*, pp. 162-167, 2006.

[2] F. Abdat, C.Maaoui, A.Pruski, "Real Time Facial Feature Points Tracking with Pyramidal Lucas-Kanade Algorithm", *Proceedings of the 17th IEEE International Symposium on Robot and Human Interactive Communication*, Munich, Germany, 2008.

[3] C. Basdogan, S. De, J. Kim, M. Muniyandi, H. Kim, M.A. Srinivasan, "Haptics in Minimally Invasive Surgical Simulation and Training", *Proceedings of the IEEE Computer Graphics and Applications*, Vol. 24 No.2, pp. 56, 2004.

[4] G. Bradski andA.Kaehler, "LearningOpenCV", *O`Reilly Media, Inc.*, 1005 Gravenstein Highway North, Sebastopol, CA 95472, 2008.

[5] J.Bouguet,"Pyramidal Implementation of the Lucas Kanade Feature Tracker Description of the Algorithm", *Intel Corporation, Microprocessor Research Lab*, 2000

[6] J. Cha, M.Eid, A.E.Saddik,"Touchable 3D Video System", ACM Transactions on Multimedia Computing, Communications and Applications, Vol.5, No.4, 2009.

[7] J. Cha, S. Kim, Y. Ho, "3D Video Player System with Haptic Interaction based on Depth Image-Based Representation", IEEE Transaction on Consumer Electronics, Vol. 52, No. 2, 2006.

[8] J. Cha, S. Kim, I. Oakley, J. Ryu, K. H. Lee,"Haptic Interaction with Depth Video Media", Advances in Multimedia Information Processing - PCM 2005, Lecture Notes in Computer Science, 2005.

[9] R. Cross, "The bounce of a ball", American Journal of Physics, Vol. 67 No. 3, 222-227 , 1999.

[10] N. Dindar, A. M. Tekalp, C. Basdogan,"Immersive Haptic Interaction With Media", Proceedings of the Visual Communications and Image Processing 2010, Huangshan, China. , 2010.

[11] V. Hayward, O. R. Astley, M. Cruz-Hernandez, D. Grant, G. Robles-De-La-Torre,"Haptic Interfaces and devices", Journal of Sensor Review, Vol. 24 No. 1, pp. 16-29, 2004.

[12] B. K. P. Horn, B.G. Schunck, "Determining Optical Flow", AI(17), No. 1-3, pp. 185-203, 1981.

[13] S. Kim, J. Cha, S. Lee, J. Ryu, Y. Ho, "3DTV System using Depth Image-Based Video in MPEG-4 Multimedia Framework", 3DTV Conference, 2007

[14] S. Kim, Y. Koike, M. Sato, "Tension Based 7 DOFs Force Feedback Device: SPIDAR-G", Transactions On Control, Automation, and Systems Engineering, Vol. 4 No. 1, pp. 9-16, 2002.

[15] X. Liu, K. Akahane, M.Isshiki, M.Sato,"Design and Implementation of an Image Haptization System", 3DSA2010(International Conference on 3D Systems and Applications）, pp.260-262, 2010.

[16] D. Morris, N. Joshi, K. Salisbury, "Haptic Battle Pong: High-Degree-of-Freedom Haptics in a Multiplayer Gaming Environment", Proceedings of Experimental Gameplay Workshop at Game Developers Conference (GDC)'04, 2004.

[17] S. O`Modhrain, I. Oakley, "Touch TV: Adding Feeling to Broadcast Media", Proceedings of the 1st European Conference on Interactive Television: from Viewers to Actors, Brighton, UK. 2003.

[18] J. Sommer, R. Arango, D. Grest, "Sparse Optical Flow for 3D Objects", 2008.

[19] M. Sato, X. Liu, J. Murayama, K. Akahane, M. Isshiki, "A Haptic Virtual Environment for Molecular Chemistry Education", Transactions on Edutainment I, LNCS 5080, pp.28-39, 2008.

[20] J. Wu, D. Wang, Y. Zhang, "Virtual Fixture based Haptic Rendering of Handwriting", IEEE International Conference on Virtual Environments, Human-Computer Interfaces, and Measurement Systems, pp.16-21, Hong Kong, China, 2009.

[21] M. Sato, "A String-based Haptic Interface －SPIDAR", ISUVR2006, Vol.191, pp.1-5,Yanjin, China, 2006