REAL-TIME GUITAR CHORD ESTIMATION BY STEREO CAMERAS FOR SUPPORTING GUITARISTS

Chutisant Kerdvibulvech   Hideo Saito

Department of Information and Computer Science, Keio University
3-14-1 Hiyoshi Kohoku-ku, Yokohama, Japan
E-mail: {chutisant, saito} @ozawa.ics.keio.ac.jp

ABSTRACT

Learning to play the guitar usually involves tedious lessons in fingering positions for the left hand. It is difficult for beginners to recognize by themselves whether they are accurately positioning their fingers on the string to make guitar chords. This paper proposes a novel approach for recognizing chords played on a guitar. In real time, it estimates whether a beginner is holding the guitar correctly. This system detects the position of a player's fingers in 3D and calculates the guitar chord which the player's left hand is producing. Stereo cameras are used to compute the 3D positions of fingers using the triangulation method. ARTag (Augmented Reality Tag) is utilized to calculate the extrinsic parameters in each frame, so that our system will allow users to move the guitar while they are playing. The experimental results have revealed that the proposed system is effective even when used in real time.

1. INTRODUCTION

A common method of learning to play the guitar involves practicing the fingering (i.e., in the left hand positioning). However, learning the guitar combines lots of elements that come together at the same time, such as holding the guitar, reading notes and the left hand finger positions.

Beginners may find it complicated to cope with all of these elements at the same time, particularly in the early lessons. While beginners are looking at the notes, they also have to pay attention to the guitar neck which their left hand is holding, and therefore it is difficult to identify whether they are holding the correct chords to produce the notes required by the musical piece that they are playing. It can be distracting for a beginner to have to read the notes and to learn the correct fingering at the same time.

In this paper, we propose an innovative guitar teaching aid that will assist guitar players by using computer vision. This system detects the players' fingers in real time and calculates the guitar chord which the player's left hand is using. We use the triangulation[5] technique to estimate 3D (3-Dimensional) positions of the player's fingers in the guitar coordinate system. We utilize ARTag[2] to compute the extrinsic parameters for estimating the guitar position as an online process. In other words, we calibrate the cameras for calculating the projection matrix in real time, and this allows the users to move the guitar's position while they are playing the guitar. As a result, beginners can recognize the guitar chords which they are playing during the song. We would consider it to be of great assistance to learners if it is able to recognize the chord played by the left hand. The system improves accuracy because it can identify whether the finger positions are correct and in accord with the finger positions required for the piece of music that they are playing (i.e., the guitar teacher inputs the correct fingering for each guitar chord and these are incorporated into the system that we have provided). As a result, beginners can immediately identify whether their fingers are in the correct position when using our system in real time. The proposed teaching system would be invaluable and have wide application for supporting guitarists to play the guitar.

2. RELATED WORK

Recently, researchers have attempted to develop musical applications of AR (Augmented Reality) as teaching aids. We found two systems that focus on teaching or assisting guitarists. Cakmakci and Berard[1] developed a system that assists beginner level musicians in learning the electric bass guitar using AR. In this electric bass guitar system, the goal is to accelerate the process of associating a musical score, the sounds represented by that score and the fingerboard. The motivation for building this system is to overcome the perceptual discontinuities that can be introduced by the dispersal of sources of information during the learning process. Motokawa and Saito[8] built a system called Online Guitar Tracking that supports a guitarist using AR. This is done by showing visual aid information on a real stringed guitar and this becomes an aid to learning to play the guitar. Online Guitar Tracking uses AR to detect the guitar so that the player can learn how to hold the strings of the guitar by overlapping the player's hand onto a manual model. Before a beginner commences a guitar lesson or is practising, he/she manually inputs the musical information (i.e., the time periods in each the guitar chord). Next, the Online Guitar Tracking displays the virtual fingers model onto the stringed guitar. Therefore, even though the user is a beginner, he/she can learn by imitating the fingering from the virtual fingers model.

However most systems examined have different goals from ours. We propose a novel supporting system for recognizing the guitar chord and estimating whether the beginner is accurately holding the chord in real time. This means that a beginner can learn more effectively how to hold the strings of the guitar when using our system.

3. PROPOSED SYSTEM

In our system, we propose an innovative approach that can estimate the guitar chord which the player's left hand is using in real time.

3.1 System Configuration

Figure 1 shows the proposed system configuration. The system is composed of two USB cameras with resolution 320x240 and a display connected to the PC for the guitar players. The two USB cameras capture the player's fingers in the left hand positioning and the guitar neck from two different positions.

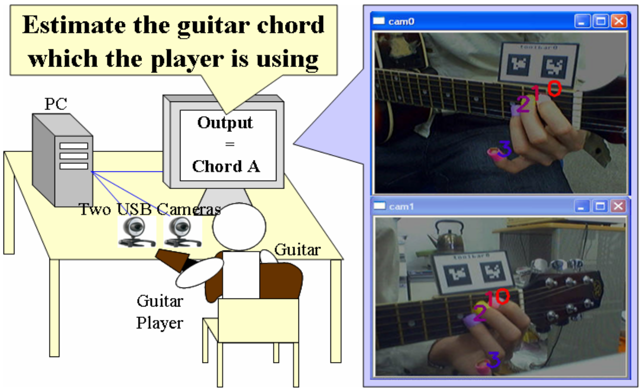

Figure 1: The proposed system configuration

The most important aspect in constructing this guitar teaching aid was developing accurate 3D positions of the fingers and the guitar. To estimate the position of the guitar from input images, we utilize an ARTag's marker in the captured image. As shown in Figure 1, we attach a small square-shaped planar marker onto the guitar and use it to compute the pose and the position of the guitar. To calculate the pose and the position of the fingers, which is directly related with the guitar position as the guitar coordinating system, we utilize four differently color markers which are attached to the four fingertips; forefinger, middle finger, ring finger and little finger respectively as shown in Figure 1. Using both the ARTag's marker and the colorful markers for tracking the guitar and fingers respectively, we can implement a practical and usable system for estimating the guitar chords in real time with just two USB cameras and a PC display.

3.2. Outline of the proposed system

Figure 2 shows the outline of the proposed system. This system is composed of four main subsystems: tracking fingers markers subsystem, real-time estimating projection matrix subsystem, calculating 3D positions subsystem and computing guitar chords subsystem. If the system tracks the pose and the position of the guitar and fingers precisely, the system can determine the guitar chord output accurately.

We will describe both the method for real-time fingers tracking and the method for real-time accurate guitar tracking that use the ARTag's marker. Firstly, the tracking fingers markers subsystem has to estimate the position of guitar player's fingers. Secondly, the real-time estimating projection matrix subsystem has the function of calculating the projection matrix in each frame. Thirdly, the calculating 3D positions subsystem has to calculate the 3D positions of the guitar and fingers. The final subsystem is a computing guitar chords subsystem which has the function of calculating the guitar chords which the guitarist is playing. The diagrams graphically describe the technique used to utilize computing the guitar chords.

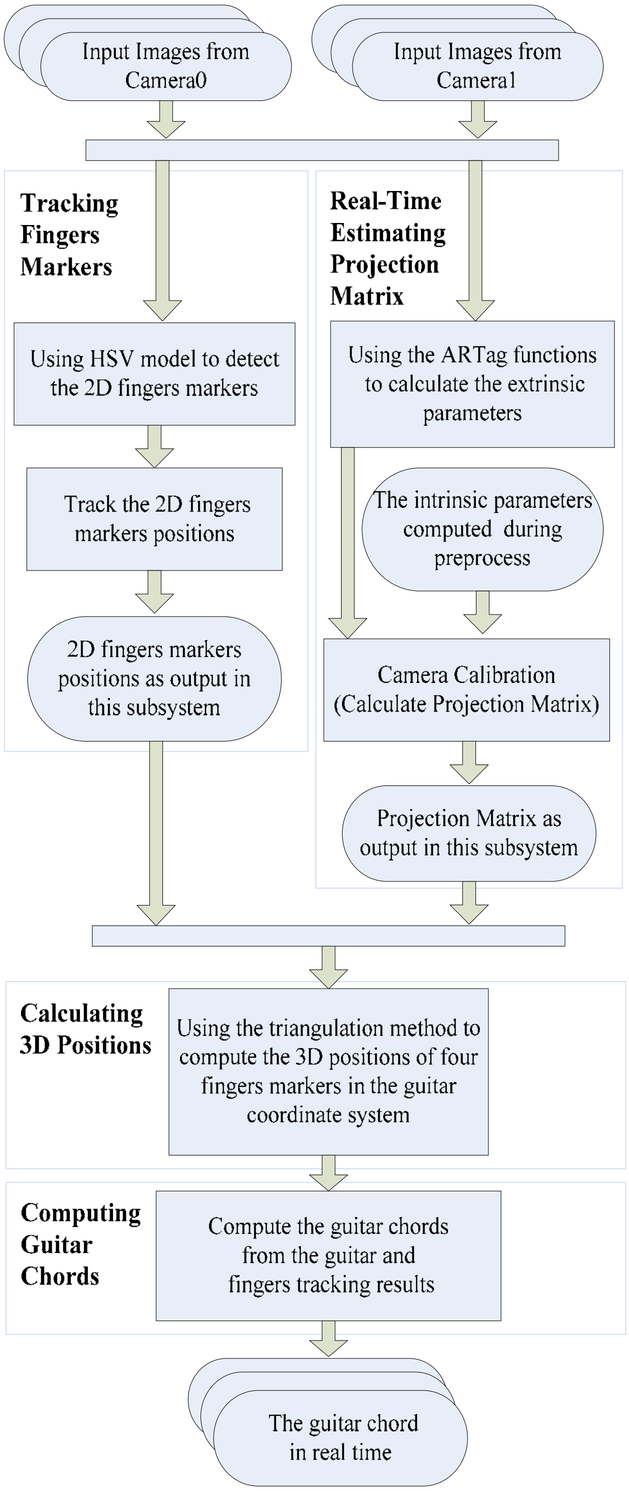

Figure 2: The outline of the proposed system

3.3 Tracking Fingers Markers subsystem

In this subsystem, we describe a method that tracks the player's fingers and estimates the pose and the position. Currently, there are many methods that present a technique to estimate the pose and the position of the fingers. For example, Wu[10] et al. proposed an advanced approach for tracking the articulated hand and fingers in a video by learning and integrating natural hand motion, so that this system can track hand articulations in a long sequence containing all the hand position information. However, to estimate the fingering required to recognize a guitar chord, it is not necessary to estimate the whole hand skeleton structure. In other words, we can estimate the guitar chord by just tracking the fingertips in real time. To track the fingertips, there are several research papers that focus on tracking fingertips using markers. For instance, Piekarski and Thomas[9] used the ARToolKit (Augmented Reality Toolkit)[6] markers which attach to the fingertips to provide three degrees of freedom in tracking the hands, and use world coordinates, to interact with a mobile outdoor AR computer. In our system, we attempted to utilize the small ARToolKit finger markers which are placed on all four fingertips; forefinger, middle finger, ring finger and little finger respectively. However, we could not successfully produce accurate finger markers by using these ARToolKit finger markers, because while the guitar player were holding their fingers, the ARToolKit finger markers in each fingertip were not entirely captured by the cameras in each frame, so that the detected results were not successful using the ARToolKit finger markers on each fingertip.

Therefore we decided to replace the finger markers from ARToolKit with colored markers. We utilize four differently colored markers which are placed on the four fingertips. We need to identify these four finger markers from background in this subsystem. The first step in the finger markers detection algorithm uses color segmentation to reject as much "non-color fingers markers" of the image as possible, since the main part of the images consists of non-color finger markers color pixels. Secondly, by using the HSV (Hue, Saturation and Value) model[4], we are able to differentiate the four color markers from the background by converting the RGB (Red, Green and Blue) picture into the HSV space. The HSV model defines a color space in terms of three constituent components: Hue, Saturation and Value. We transform RGB space to HSV coordinate system and find the proper threshold values of Hue, Saturation and Value in each marker. Thus, we are able to differentiate the four color markers from background, and therefore we are able to track the movement of the player's fingers by just using the four color markers on each fingertip.

3.4 Real-Time Estimating Projection Matrix subsystem

In this subsystem, we describe the method we used to calculate the projection matrix in every frame. To estimate 3D position of finger at online process, it is necessary to compute the projection matrix relative to the world coordinate that is attached to the guitar neck. The world coordinate is defined on the top right corner of guitar neck as shown in Figure 3. Because the guitar neck always moves while the cameras are fixed, the projection matrix also changed at every frame.

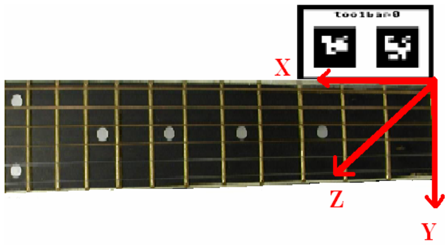

Figure 3: The world coordinate on the top right corner of guitar neck

As shown in Figure 3, there is an ARTag's marker on the guitar. Marker tracking is achieved using the publicly available ARTag functions. To estimate the pose and the position of the guitar marker in 3D space, it is necessary to correctly measure the positions, orientations and focal lengths of the two cameras. This is called Camera Calibration[3]. In camera calibration process, the relation by projection matrix is generally employed as the method of describing the relation between the 3D space and the images. To calculate projection matrix, it is necessary to calculate two types of camera parameters: intrinsic parameters and extrinsic parameters as following equation

$$P = A[R,t] = \begin{bmatrix} \alpha_u & -\alpha_u \cot\theta & u_0 \\ 0 & \alpha_v / \sin\theta & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} R_{11} & R_{21} & R_{31} & t_x \\ R_{12} & R_{22} & R_{32} & t_y \\ R_{13} & R_{23} & R_{33} & t_z \end{bmatrix} (eq.1)$$

Projection Matrix P → Intrinsic Matrix A
Extrinsic Matrix and Vector R , t

The important camera properties, namely the intrinsic parameters that must be measured, include the center point of the camera image, the lens distortion and the camera focal length. We compute the intrinsic parameters during preprocess the first time. As shown in Equation 1, the matrices R and t in the camera calibration matrix describe the position and orientation of the camera with respect to world coordinate system. The extrinsic parameters include three parameters for the rotation, and another three for the translation. Using the online process, the ARTag function computes the extrinsic parameters in every frame, and therefore we can compute in real time the projection matrix from both the intrinsic parameters and the extrinsic parameters using Equation 1. It is necessary to calibrate the cameras in every frame because our system allows the users to move the guitar's position when they are playing the guitar.

3.5 Calculating 3D Positions subsystem

In this subsystem, we utilize the triangulation to compute 3D position of fingers in the guitar coordinate system. Figure 4 depicts the triangulation method which demonstrates how to estimate 3D positions from two cameras using the projection matrix results. While **m**(u,v) is defined in the first of the image coordinate, **m**'(u',v') is defined in the second image coordinate system. **M**(X,Y,Z) is defined in the world coordinate system.

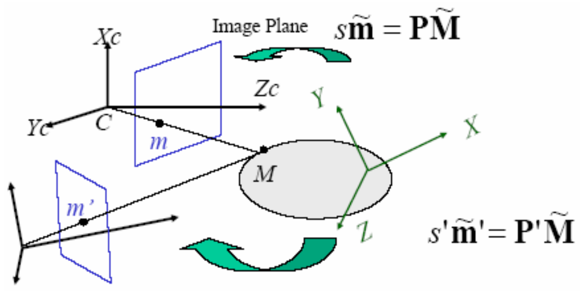

Figure 4: The triangulation method which represents how to estimate 3D positions from two cameras

The 3D positions in the world coordinate system can be estimated using the projection matrix results as shown in Equation 2. In this way, using the data gathered from the finger detection and the projection matrix results, we can estimate the 3D position of the finger markers in the guitar coordinate system in real time. Because we are able to estimate the 3D position in the guitar coordinate system, it allows the users to move the guitar's position while they are playing. The results of tracking 3D positions will be successful because we compute both of the guitar position and fingers positions based on the guitar coordinate system.

$$\begin{bmatrix} P_{14} - uP_{34} \\ P_{24} - vP_{34} \\ P'_{14} - u'P'_{34} \\ P'_{24} - v'P'_{34} \end{bmatrix} = \begin{bmatrix} uP_{31} - P_{11} & uP_{32} - P_{12} & uP_{33} - P_{13} \\ vP_{31} - P_{21} & vP_{32} - P_{22} & vP_{33} - P_{23} \\ u'P'_{31} - P'_{11} & u'P'_{32} - P'_{12} & u'P'_{33} - P'_{13} \\ v'P'_{31} - P'_{21} & v'P'_{32} - P'_{22} & v'P'_{33} - P'_{23} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} (eq.2)$$

3.6 Computing Guitar Chords subsystem

We describe the method to compute the guitar chords from the guitar and fingers tracking results in this subsystem. It is necessary to classify each guitar chord, so that we can estimate the guitar chord at the online process. Moreover, in this subsystem, we describe the indirect way to solve the Forefinger's Pushing Problem too.

3.6.1 Estimate each guitar chord

We need to classify each guitar chord, and therefore we utilized the interval threshold technique to estimate the guitar chord in real time. For classification data of each guitar chord, we collected the required information (i.e., the interval threshold 3D positions of four finger markers) of every guitar chord by experimenting. In other words, we decide the interval threshold positions from 100 frames of each guitar chord by playing a sample of the notes. At every finger marker, we select both of the minimum value and the maximum value of the 3D positions from these 100 frames, and then we utilize these minimum and maximum values as the interval threshold positions of each fingers marker. Once the threshold positions are correctly identified, we can use these threshold positions to classify each guitar chord. We called these interval threshold positions the "Guitar Chord Database" in each guitar chord. The information of the Guitar Chord Database in each guitar chord consists of eight main values as shown in Table 1.

Table 1 depicts the structure of Guitar Chord Database which explains how it collects the datum in each guitar chord. In each finger marker, the Guitar Chord Database gathers datum both of the minimum and the maximum 3D positions values. Next, in each 3D positions value, it composed of three sub-values in each axis: x axis, y axis and z axis respectively.

	Minimum positions values of each axis	Maximum positions values of each axis
Fore-finger marker	MinForeFingX MinForeFingY MinForeFingZ	MaxForeFingX MaxForeFingY MaxForeFingZ
Middle finger marker	MinMiddleFingX MinMiddleFingY MinMiddleFingZ	MaxMiddleFingX MaxMiddleFingY MaxMiddleFingZ
Ring finger marker	MinRingFingX MinRingFingY MinRingFingZ	MaxRingFingX MaxRingFingY MaxRingFingZ
Little finger marker	MinLittleFingX MinLittleFingY MinLittleFingZ	MaxLittleFingX MaxLittleFingY MaxLittleFingZ

Table 1: The structure of Guitar Chord Database in each guitar chord

Once the player has played the guitar, our system would analyze the inputted tracking results (i.e., the inputted 3D position of the four fingers markers which we need to identify the chord) in each frame. Next, we use both of our inputted tracking results and the Guitar Chord Database to estimate the guitar chord at online process. In other words, while the inputted tracking results are computed, our system compares these inputted tracking results with Guitar Chord Database datum to identify the guitar chord, so that our system can immediately recognize the detected guitar chord in real time. As a result, once a player has played the guitar, our system can compute which the guitar chord is being played in real time.

3.6.2 Forefinger's Pushing Problem

In the real guitar playing, there are some guitar chords which require the guitarists to push all six strings down with the one finger (the so called bar chord). This usually only involves the forefinger. For example, as shown in Figure 5(a), a six stringed line staff that represents the guitar fret board indicates which string and frets are played. We called the six stringed line namely the e string, the B string, the G string, the D string, the A string and the E string ordered from the top down accordingly. It is obvious that the guitar chord F, F# and Cm required the guitarists to push all six strings down with the forefinger in the 1st, 2nd and 3rd frets respectively.

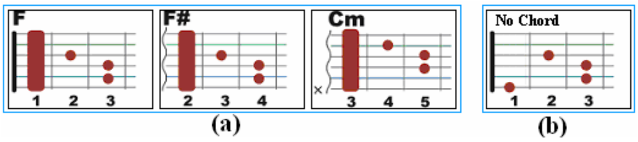

Figure 5: The sample of guitar chord charts: (a) The samples of guitar chord charts which have a Forefinger Pushing Problem; (b) The guitar chord chart which does not really exist in all real guitar chords (i.e., the forefinger position is on the string E which is the same fret as the guitar F chord, while the other three fingers push exactly the same frets and string as the F chord)

In our method, we utilize only four color markers which are placed on the four fingertips, so that it could difficult to identify whether the forefinger pushed all six strings down. We called this the Forefinger's Pushing Problem.

However, we do not have to solve this problem directly because there are no two guitar chord in all the guitar chords when the three fingers (i.e., the middle finger, the ring finger and the little finger) push exactly the same frets and strings, however the forefinger from first chord pushes all six strings down while the forefinger from second chord pushes the E string on the same fret as the first chord. In other words, we would classify each guitar chord to make it different by only using the four fingers markers.

For instance, in the F chord as shown in Figure 5(a), the forefinger is required to push all six strings down on the first fret, while the middle finger, the ring finger and the little finger are pushed in the G string (2^{nd} fret), the D string (3^{rd} fret) and the A string (3^{rd} fret) respectively. In this way, there is no other chord in all the 144 guitar chords (i.e., this information came from [7]) in which the forefinger's position pushes the E string (1^{st} fret) while the fret and string positions in the other three finger's positions are exactly the same as the F chord (i.e., the middle finger, the ring finger and the little finger are pushed in the string G (2^{nd} fret), the D string (3^{rd} fret) and the A string (3^{rd} fret) sequentially) as shown in Figure 5(b), so that the information of the four fingertips can differentiate each chord by using only four fingertip markers. As a result, there are no two chords which we can not differentiate using our method, and therefore we can classify all guitar chords with just four finger markers. It means that we can successfully solve this Forefinger's Pushing Problem using indirect method.

4. EXPERIMENTAL RESULTS

In order to show the efficiency of our method, we evaluate accuracy and stability of our system by testing 15 samples guitar chords. Figure 6 shows the scene that a guitar player uses our system. As shown in Figure 6(a), the cam0 and cam1 windows depict the input images which are captured from two USB cameras. These two USB cameras capture the player's fingers in the left hand positioning in real time. In Figure 6(a), the guitar player is holding the chord A, and then the four color circles in each fingertip depict four 2D tracking results of fingers markers. In Figure 6(b), the 3D reconstruction window, which is drawn by OpenGL, represents both of the detected 3D positions of the four finger markers and the guitar chord output. The four small differently color cubes show each detected result of the finger markers in 3D space, and then the guitar chord output represents the detected chord in the guitar coordinate system. As shown in the 3D reconstruction window in Figure 6(b), it is obvious that our system can estimate the guitar chord A correctly which is the same chord as the player is holding.

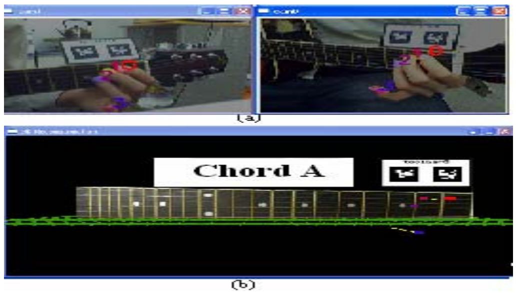

Figure 6: The sample of experimental result: chord A**:** (a) The input images are captured from two cameras, holding chord A**;** (b) The detected 3D positions of four fingers markers and the guitar chord A output

To classify each guitar chord, it is quite difficult to obtain the best results because the four fingering positions in each chord are not markedly different. However, from testing the 15 samples guitar chords in our experiments as shown in the Table 2, it is generally observed that our system can estimate the 12 guitar chords (i.e., chord D, D7, Dm, E, Em, G, G7, F#m, F, Cadd9, Cm and F#) correctly with over 75 percentages, while the other 2 chords (i.e., chord A and Am) can be estimated accurately with over 50 percentages. The experimental results revealed that the proposed system is effective even when used in real time.

ESTIMATE of CHORDS	FACT of CHORDS (%)														
	A	Am	C	D	D7	Dm	E	Em	G	G7	F#m	F	Cadd9	Cm	F#
A	59	-	-	-	-	-	-	-	41	-	-	-	-	-	-
Am	-	51	-	14	35	-	-	-	-	-	-	-	-	-	-
C	-	-	25	28	-	21	-	-	26	-	-	-	-	-	-
D	-	-	-	100	-	-	-	-	-	-	-	-	-	-	-
D7	-	-	-	-	81	19	-	-	-	-	-	-	-	-	-
Dm	-	-	-	-	-	100	-	-	-	-	-	-	-	-	-
E	-	12	-	-	-	12	76	-	-	-	-	-	-	-	-
Em	-	-	-	-	-	10	12	78	-	-	-	-	-	-	-
G	-	-	-	-	-	-	-	-	92	-	-	-	8	-	-
G7	-	-	-	-	-	-	-	-	-	86	14	-	-	-	-
F#m	-	-	-	-	-	-	-	-	-	-	93	-	-	-	7
F	-	-	-	-	-	-	-	-	-	-	-	100	-	-	-
Cadd9	-	-	-	-	-	-	-	-	-	-	17	-	83	-	-
Cm	-	-	-	-	-	-	-	-	-	-	-	7	-	93	-
F#	-	-	-	-	-	-	-	-	-	-	4	-	-	-	96

Table 2: Evaluation of the proposed method by testing 15 samples guitar chords from our experimental results: The vertical columns represent the fact of guitar chords which the player is holding while the horizontal rows depict the estimation of guitar chords using the proposed method. The numbers in this table show the estimated experimental results from 15 samples chords in units of percentages.

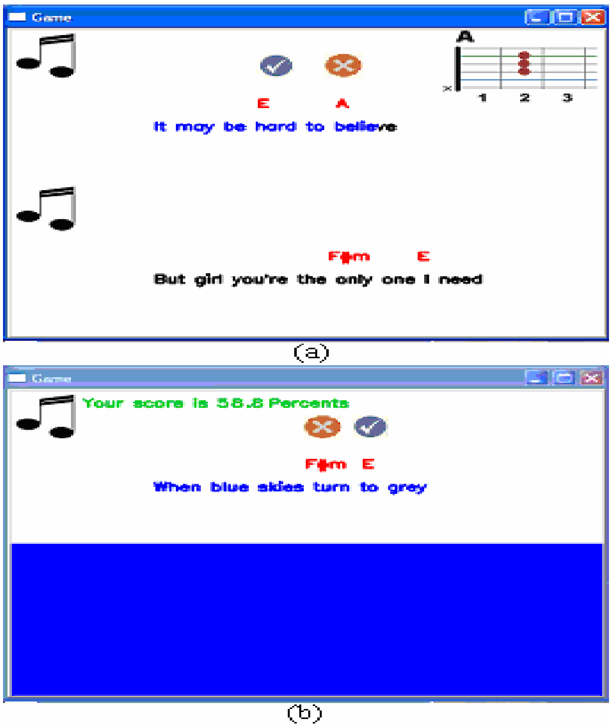

Figure 7: The user interface of the novel guitar application applying our method: (a) The guitar application which represents song lyrics (i.e., the lyrics are displayed, using color changes that are synchronized with the music), guitar chord charts (i.e., as shown in the top right of this figure to guide the guitarists to identify how the player's left hand should hold each chord), and the users evaluation as to whether they are correctly holding each chord (i.e., as shown in the small right and wrong figures above the chords); (b) The guitar application produces an overall evaluation score of the users' accuracy automatically when the song is finished. This is displayed in percentage units as shown in the green characters on the top of this figure.

Furthermore, after we recognize the guitar chord which the player's left hand is using, we apply the proposed method to utilize this novel guitar application. Figure 7 represents the innovative guitar application for recognizing the guitar chord whenever the player is holding it correctly by applying the proposed method. This guitar application contains the lyrics, guitar chord charts and voice information that can be used to assist student guitarists to more effectively learn to play a song. The lyrics are displayed on the screen in color which changes and is synchronized with the music. Moreover, this application recognizes each successive chord used by the song, whenever the player is holding it correctly and this provides greater user friendliness. It is considered that this would be of great assistance to guitarists because they are able to immediately identify if their own finger positions are correct and if they are in accord with the correct chords required by the musical piece and all the information is provided in real time. In addition, this application will show an overall evaluation score, in units of percentages, indicating the user's accuracy when the performance has been completed. This method would be invaluable as a teaching aide for guitar players.

5. CONCLUSIONS

In this paper, we propose an effective teaching tool for learning to play the guitar using computer vision in real time.

We described the method to achieve the system. We use ARTag's marker information and color markers to track the guitar and the correct fingering respectively. We utilize the results of the guitar and the finger detection to compute the guitar chord in real time. As a result, it would be of assistance to the guitarists by providing real time feedback by checking it against the preloaded accurate finger positions required by each musical piece. Applying our method, it would allow the player to gain a higher level of enjoyment during the lesson and at the same time provide an accelerated learning method.

We believe that we can successfully produce an accurate system output. However, we are planning to further refine the problem of the ARTag marker. We look forward to making technical improvements to our system that may involve replacing the markers with computer vision algorithms which may result in even providing even greater user friendliness.

6. REFERENCES

[1] O. Cakmakci and F. Berard. "An Augmented Reality Based Learning Assistant for Electric Bass Guitar," In Proc. of the 10th International Conference on Human-Computer Interaction, 2003.

[2] M. Fiala. "ARTag Revision 1: A Fiducial Marker System Using Digital Techniques," Published as NRC/ERB-1117, NRC Publication Number: NRC 47419, 2004.

[3] D. A. Forsyth and J. Ponce. "Computer Vision: A Modern Approach," Prentice Hall, ISBN 0-130-85198-1, 2002.

[4] R. Gonzalez and R. E. Woods. "Digital Image Processing," 2 ed, Prentice Hall Press. p. 295, ISBN 0-201-18075-8, 2002.

[5] R. Hartley and A. Zisserman. "Multiple View Geometry in Computer Vision," 2 ed, Cambridge University Press, ISBN 0-521-62304-9, 2004.

[6] H. Kato and M. Billinghurst. "Marker Tracking and HMD Calibration for a Video-based Augmented Reality Conferencing System," In Proc. of the 2nd International Workshop on Augmented Reality, pp. 85–94, 1999.

[7] C. Lopez. "The "First Stage" Guitar Chord Chart - Learn How To Play The Most Commonly Played Guitar Chords," Publisher: First Stage Concepts, ISBN 0-966-77191-5, 2006.

[8] Y. Motokawa and H. Saito. "Support System for Guitar Playing using Augmented Reality Display," In Proc. of the 5th IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR '06), 2006.

[9] W. Piekarski and B. H. Thomas. "Using ARToolKit for 3D Hand Position Tracking in Mobile Outdoor Environments," In ART02, 1st International Augmented Reality Toolkit Workshop IEEE, 2002.

[10] Y. Wu, J. Lin and T. S. Huang. "Analyzing and Capturing Articulated Hand Motion in Image Sequences," IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 27, No. 12, 2005.