COMPUTER VISION METHOD FOR GUITARIST FINGERING RETRIEVAL

Anne-Marie Burns, Marcelo M. Wanderley
McGill University
Schulich School of Music

ABSTRACT

This article presents a method to visually detect and recognize fingering gestures of the left hand of a guitarist. This method has been developed following preliminary manual and automated analysis of video recording of a guitarist. These analysis permitted to determine some important characteristics of a vision system for guitarist fingering, namely the focus on the effective gesture, the consideration of the action of each individual finger, and the definition of a recognition system not relying on comparison against a knowledge base of previously learned fingering positions. Motivated by these results, studies on three important aspects of a complete fingering system were conducted. One study was on finger tracking, another on strings and frets detection, and the last one on movement segmentation. Finally, a prototype that integrated the above three aspects in a system for guitarist left hand fingering detection has been developed.

1. INTRODUCTION

Fingering is an important aspect of the sound production in guitar playing especially because in the case of a fretted instrument many possible combinations of string, fret, and finger can produce the same pitch. Fingering retrieval is an important topic in music theory, music education, automatic music generation and physical modeling. Unfortunately, as Gilardino [6] [7] noted, fingering information is rarely present on score indications.

Fingering can be retrieved at different moments of the music production process. Three main strategies are:

- Pre-processing using score analysis;
- Real-time using Midi guitars;
- Post-processing using sound analysis;

Radicioni, Anselma, and Lombardo [14] retrieve the fingering information from the analysis of the score. The score is fragmented in phrases and the optimum fingering for each phrase is determined by finding the shortest path in an acyclic graph of all possible fingering positions. Weights are assigned to each position based on a set of rules. The problem of this approach is that it cannot account for all the factors influencing the choice of a specific fingering, namely philological analysis (interpretation of a sequence of notes), physical constraints due to the musical instrument, and biomechanical constraints in the musician-instrument interaction. Outputs of these systems are similar to human solutions in many cases but can hardly deal with situations where the musical intent has more importance than the biomechanical optimum fingering.

Other systems retrieve the fingering during or after a human has played the piece. One of these approaches uses a Midi guitar. Theoretically, using a Midi guitar with a Midi channel assigned to each string, it is possible to know in real-time what pitch was played on what string and consequently the fret position can be determined. In practice, Midi guitar users report several problems including the variation in the recognition time from one string to another and the necessity to adapt their playing technique in order not to create glitches or false note triggers [17].

An approch using the third strategy is the study of the guitar timbre. Traube [15] suggested a method relying on the recording of a guitarist. The method consists of analyzing the sound to identify the pitch, find the plucking point and then determine the string length to evaluate the fingering point. Shortcomings of this method are that actually it cannot be applied in real time, it works only when one note is played at the time, and the error range of the string length evaluation, therefore the fingering evaluation, is better than one centimeter in the case of open strings but can go up to height centimeters in the case of fretted strings [16].

This paper presents an alternative method for real-time retrieval of the fingering information from a guitarist playing a musical excerpt. It relies on computer analysis of a video recording of the left hand of the guitarist. The computer treatment is inspired by the research on tabletop applications [10][11] and by the research on 3-dimensional finger-tracking [8]. These applications often use specific and expensive hardware (infrared camera for example) and limit the tracking to one finger. In this paper we suggest an alternative method that can work with simple hardware, such as a low-cost webcam, and that extends the tracking to all fingertips.

The first part of this article is a discussion about the preliminary manual and automated analysis of multiple views recordings of a guitarist playing a variety of musical excerpts. The theoretical fundation of the Hough transform algorithm used in the following sections is then explained. The subsequent sections present studies on three important aspects of visual analysis of a guitarist fingering, namely finger tracking, strings and frets detection and movement segmentation. Finally a system integrating these three components is presented.

2. PRELIMINARY ANALYSIS

During the preliminary analysis, different camera views were evaluated (global view, front view, and top view). The aim was to find a viewpoint that allows the retrieval of the desired information with the desired degree of accuracy and precision. As it will be shown, these two objectives are conflicting. Accuracy and precision necessitate a close viewpoint focusing on one point of interest, therefore, loosing complementary information. These analysis were based on the hypothesis that the computer would be able to see at least what the human eyes can see, i.e. they were based on human observations of the images captured by the camera.

2.1. Global View

As it can be observed in figure 1(a), the global view is ideal for its richness in gestural information. This view permits to see the overall posture of the guitarist and also to see the action of both hands on the guitar neck and near the sound hole. This view is also rich in information about the expressive contents since the face can be observed. Unfortunately, using this view it is impossible to obtain a detailed image of the hands (e.g. fingering or plucking information). To solve the fingering problem, a close-up on the neck region is necessary.

2.2. Front View

By focusing on the left hand as seen in figure 1(b), it is possible to obtain a more detailed image of the neck. Of course, using this view, right hand, postural and facial gesture information are lost. On the other side, this view provides a clear vision of the fingers in most of the situations, although some occlusion may happen with specific finger positions. Frets and strings are also visible and consequently could be detected to help the estimation of the finger position on the neck. However, a drawback of this view is that it is not possible to determine if a string is pressed or not.

2.3. Top View

Figure 1(c) presents a different perspective on the region observed with the front view. This view presents characteristics similar to the front view, namely a detail view of the fingers, the possibility to detect strings and frets and the potential occurrence of the finger occlusion problem. Moreover, this view permits to observe the fingers proximity to the string; it may therefore be possible to know if the string is pressed or not by the guitarist. Another potential interest of this view is that it is close to the view the musician has of the neck when playing. This may or not have influence in the computer system but it could be interesting, perhaps, in a system design for educational purposes.

(a) Global view with zooms

(b) Front view of the left hand

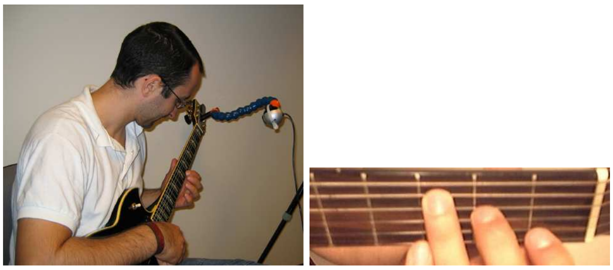
(c) Top view of the left hand

Figure 1. Three different views of a guitarist playing captured from a camera on a tripod placed in front of the musician: (a) Global view with zoom on different important zones for gesture analysis, namelly facial expression and front view of the left and right hand. (b) Front view of the left hand. (c) Top view of the left hand.

2.4. Top View with the Camera Mounted on the Guitar

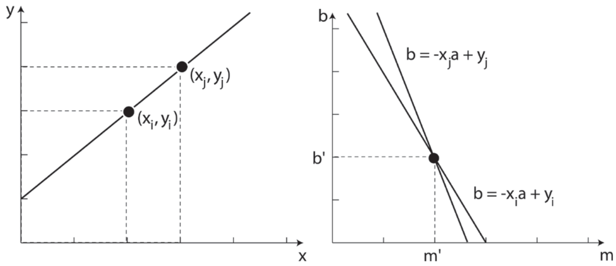
The top view (figure 1(c)) was retained for its interesting characteristics with respect to the problem, namely a detailed view of the fingers, the possibility to detect strings and frets, and the possibility to observe the finger-string proximity. However, slow motion observations of the video recording showed that the neck is subject to a lot of ancillary movements. Notably, three-dimensional rotational movements could be observed. These movements result in an unstable position of the neck in the image. The neck to camera position is subject to change in distance, view-angle, and the captured neck portion varies due to translational movements. Preliminary automated tests have shown that this can influence the computer capacity to correctly identify the fingering. Consequently, the camera on a tripod was replace by a camera mount on the guitar neck (figure 2).

The preliminary automated fingering recognition tests were performed by comparing two top view recordings of a musician playing the same excerpt. Eleven chords of the first recording were chosen and stored in a knowledge base in the form of Hu moments vectors [12]. These chords were compared with stable portions of the second



(a) Camera mount (b) Camera view

Figure 2. This figure illustrates the guitar mount that was used to eliminate the ancillary gesture problem: (a) Picture of the camera mount installed on an electric guitar. (b) View of the camera on a classical guitar. In this picture, the camera is set to capture the five first frets



(a) Collinear points (b) Concurrent lines

Figure 3. Parametrization using the line equation: (a) Two collinear points in the xy plane; (b) Intersection in the mb plane of the concurrent lines representing the points i and j.

recording. Stable portions selection was done by performing movement segmentation. In other words, the transitory phases of preparation and retraction were eliminated based on the assumption that they are high motion phases, and the nucleus phase was retained. These preliminary tests permitted to identify some difficulties of the guitarist fingering problem:

1. Using an appearance base method limits the system to previously learned material.
2. Using the global shape of the hand limits the system to the recognition of chords.
3. Using a knowledge base makes the recognition time grow with the knowledge base size.

In conclusion, these preliminary analysis permitted to get insights on how to visually resolve the guitar fingering problem. Therefore, the main specifications for a fingering recognition system are:

1. Focus on effective gestures by further reducing the presence of ancillary gestures and background elements.
2. The use of a representation that more precisely considers the action of individual fingers.
3. The use of a recognition mechanism that eliminate the burden of a knowledge base and that is therefore not limited to previously learned material.

The first specification can be achieved using the guitar mount as presented in figure 2. In order to fulfill the other specifications, three studies were conducted. In a first study, the circular Hough transform was chosen to perform finger-tracking. A second study ported on strings and frets detection and a third one on movement segmentation.

3. THE HOUGH TRANSFORM THEORY

The Hough transform is an important concept in pattern matching. It uses the mathematical description of a geometric shape to find regions of an image that best fits that shape. Its use in computer vision is born from the observation that industrial and natural images contain shapes that can be approximate by geometric shapes. In this paper, two kind of Hough transform will be used:

1. The linear Hough transform will be use to detect the guitars strings and frets;
2. The circular Hough transform will be use to detect fingertips which end can be estimate by a semi-circular shape.

3.1. The linear Hough Transform

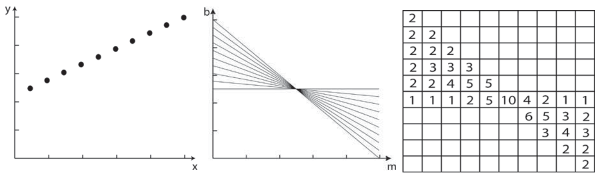
The original method proposed by Hough [9] is a simple but efficient one. Hough first considered the line equation:

$$y_i = mx_i + b \quad (1)$$

and observed that even if infinitely many lines pass through (x_i, y_i) they all satisfy equation 1. He therefore modified the equation to work in the mb plane:

$$b = -x_i a + y_i \quad (2)$$

Using this parameters space all points contained on a line with slope m' and intercept b' will intersect at (m', b'). This fact is illustrated in figure 3. In the real domain there exist infinitely many lines that pass through a point, the mb parameters space therefore needed to be discretized. This is done by dividing the parameters space in accumulator cells. Each point is then tested for all possible m value of the discrete mb space. The cells are called accumulators because they are incremented each time a point is tested on their (m, b) coordinates. Maxima in the mb space correspond to detected lines in the xy space. Figure 4 gives an example of ten points that can be link into a line. Figure 4(b) displays the solutions for the line in fig:linediscrete-a. It can be observed that to collinear points in the xy space corresponds concurrent lines in the mb space. These lines intersect at the (m, b) coordinates corresponding to the line that best fit the collinear points. The cell corresponding to the intersection contains a maximum as it can be observed in figure fig:linediscrete-c. One problem with this approach is that the intercept and the



(a) Collinear points (b) Concurrent lines (c) Accumulator cells

Figure 4. Example of 10 collinear points: (a) Collinear points; (b) Concurrent lines; (c) Accumulator cells.

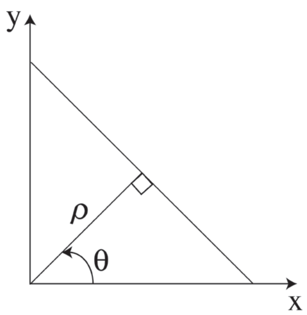
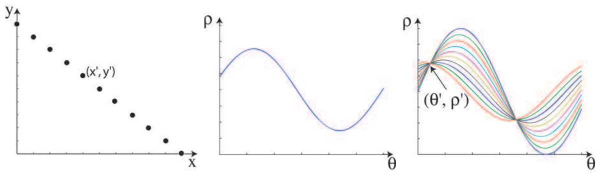


Figure 5. Normal parametrization of a line given by equation 3 where θ and ρ are fixed parameters

slope approach infinity as the line approaches the vertical. This problem can be solved by using the normal representation of a line, as explained by Duda and Hart [4]. As illustrated in figure 5, a line can be described by the angle θ of its normal and by its distance ρ from the origin using the following equation:

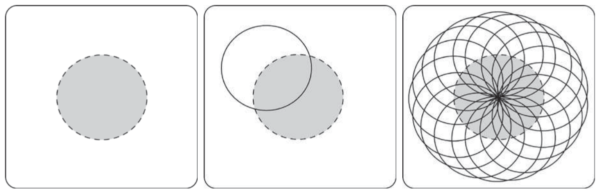
$$x \cos\theta + y \sin\theta = \rho \qquad (3)$$

In this method, each point will be tested on a discrete interval of θ comprise between $[0, \pi]$ as it was tested for all possible m in the previous case. Figure 6 shows that points on a same line will have concurrent curves in the (θ, ρ) space. As it was the case in the (m, b) space, the problem is therefore to find these intersect points that are represented by maxima in the two-dimensional accumulator cells array.



(a) Collinear points (b) One point in the (θ, ρ) space (c) The 10 points

Figure 6. The normal parametrization of the line: (a) Ten collinear points; (b) ρ as a function of θ, applying equation 3 for fixed parameter x and y; (c) Intersection of the concurrent curves representing the ten points in a.



(a) A circular shape (b) Hough transform apply on (a) (c) Maxima at the center of (a)

Figure 7. Circular Hough Transform: (a) A circular shape; (b) Circles of a given radius are drawn along the contour of the circular shape; (c) The intersection of all circles indicate the center of the detected circular shape.

3.2. The circular Hough Transform

The concept under the circular Hough transform is rather the same as the linear one and was introduced by Duda and Hart in 1972 [4]. As they mention in their article, the Hough transform can be applied on any shape that can be represented by a parametric equation. The number of dimensions, and therefore, the complexity of the accumulator cells array depends on the number of parameters.

In the case of the circle, equation 4 can be used and would generate a three-dimensional array with parameters space (a, b, r) where (a, b) are the coordinates of the center of the circle and r is the radius.

$$(x - a)^2 + (y - b)^2 = r^2 \qquad (4)$$

When it is possible, it is advantageous to reduce the parameters space to only (a, b) and to test the image over a fixed radius or a reduced set of r. In this case, equation 6 can be used.

$$x = a + r \sin\theta \qquad (5)$$

$$y = b + r \cos\theta \qquad (6)$$

Figure 7 illustrates how the Hough transform is applied to the contour image of a circular shape. Figure 7(a) represents the circular shape to detect. Figure 7(b) demonstrates how circles of a given radius are drawn around the contour image. Discrete points of the circular shape contour are used as center (the circular contour is shaded for clarity). Finally, figure 7(c) shows that in the case of a match (circular shape of the search radius) all drawn circles will intersect at the center of the detected circle. This will translate in a maximum in the accumulator cells array.

4. FINGER TRACKING

The circular Hough transform algorithm used in this paper was developed and tested in EyesWeb at the InfoMus laboratory during a study on the development and evaluation of diverse finger-tracking algorithms [2]. Between all the tested algorithms, the circular Hough transform was retain for the guitarist fingering problem due to its interesting characteristics:

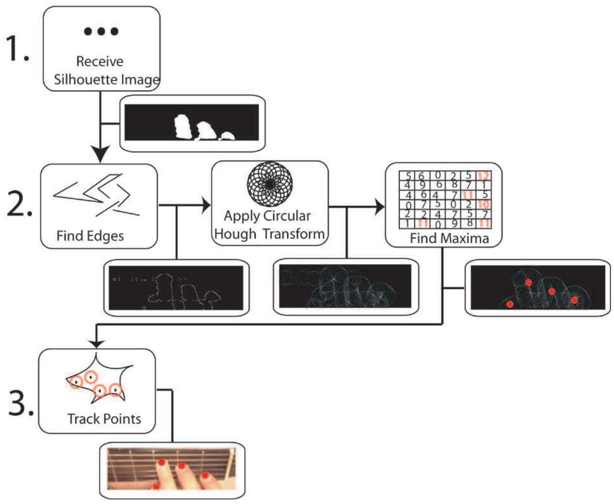


Figure 8. Fingertips detection using the circular Hough transform algorithm

1. It demonstrated to have a high degree of precision and accuracy;
2. It can be applied in complex environments and with partial view of the hand;
3. It can work on edge version of the images.

4.0.1. Circular Hough Transform

As illustrated in figure 8, the circular Hough transform is applied on the binary silhouette image of the hand. The edge image is obtained by applying the Canny edge detection algorithm [3] on the silhouette images. The circular Hough transform algorithm uses the fact that the finger ends have a quasi-circular shape while the rest of the hand is more linearly shaped. In this algorithm, circles of a given radius are traced on the edge images and regions with the highest match (many circles intersecting) are assumed to correspond to finger ends' center.

5. STRINGS AND FRETS DETECTION

By tracking the fingertips it is possible to know where each finger is in space. In the case of fingering on the guitar, this space can be defined in terms of strings and frets coordinates. The detection of strings and frets in the image is consequently a determinant step. Figure 9 shows the strings and frets detection algorithm. Prior to strings and frets detection, the region of interest, in that case the guitar neck, must be located in the image. Once the neck has been located, the strings and frets are segmented from the gray-scale neck image by applying a threshold. A vertical and a horizontal Sobel filter are applied on the threshold image in order to accentuate the vertical and horizontal gradients. A Linear Hough Transform is then computed on the two Sobel images. The linear Hough transform detects linearity in group of pixels therefore creating lines. These lines are then grouped by proximity in order to determine the position of the six strings and of the frets.

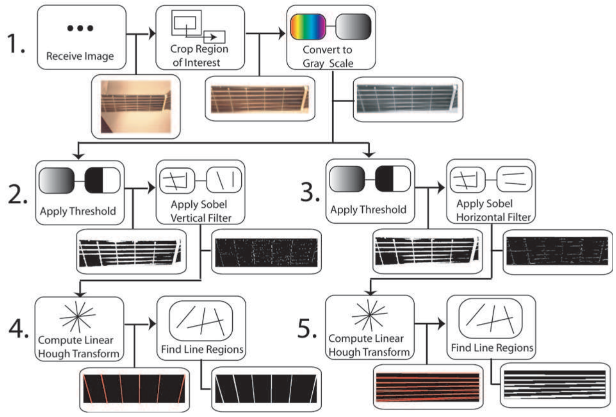


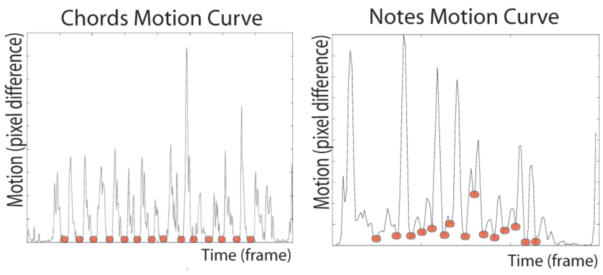
Figure 9. Strings and Frets detection using the linear Hough transform algorithm

Once this is done, it is possible to create a grid of coordinates to which fingertip positions can be matched.

6. MOVEMENT SEGMENTATION

Movement segmentation is essential in order to detect fingering positions during the playing sequence. Furthermore, in order to save computer resources, this segmentation is done early in the algorithm so that the subsequent analysis steps are performed only on significative finger positions (see figure 11 line 3). Movement segmentation is used to separate the nucleus part of the gesture from the preparation and retraction parts. Assuming that the temporal division of empty-handed gestures in three phases (preparation, nucleus, retraction) is correct and consistent [13], a similar structure can be used to analyze instrumental gestures.

In the preliminary analysis, movement segmentation was done by applying a threshold on the motion curve (figure 10 a) generated by the computation of the pixel difference between each frame. The characteristic lower velocity phase of the nucleus was easily detected between each chord. In other playing situations, however, like when playing a series of notes, the separation between the movement transitory phases and the nucleus is not that clear (figure 10 b). This is due to a phenomenon called anticipatory placements of action-fingers that has been studied in violin [1] and piano [5]. In these cases, the preparation phase of other fingers happens during the nucleus of the action-finger. The motion is therefore not serial and consequently, the motion curve does not exhibit clear global minima like in the case of plucked chords. However, local minima can still be observed and detected and are assumed to correspond to the moment the note is trigged by the right hand. Actually, the prototypes have no mechanisms to prove this assumption but the phenomenon is observable in all the preliminary data and is consistent with other instruments bimanual synchronization movement studied in the literature [1]. Local minima are found by computing the second derivative of the motion curve.



(a) Chords motion curve (b) Notes motion curve

Figure 10. (a) Motion curve of a guitarist playing chords (b) Motion curve of a guitarist playing notes

Since the prototypes work in real-time with discrete data, that is done by subtracting the signal with its delayed version twice.

7. PROTOTYPE

The prototype was designed to fulfill the requirements for a fingering recognition system highlighted by the preliminary analysis. The focus on effective gestures was partially realized at hardware level by attaching the camera to the guitar neck thereby eliminating the motion of the neck caused by the guitarist ancillary motion. Elimination of the background elements is done by the selection of a strict ROI (Region of Interest) around the neck and by applying a background subtraction algorithm to the image. Movement segmentation is performed by finding minima in the motion curve obtained by computing the difference of pixel between each frame as explained previously. The action of each individual finger is considered using the finger-tracking algorithm described above. This prototype recognizes gestures by matching the fingertip positions to the strings and frets grid of coordinates and therefore does not rely on any knowledge base. The details of the algorithm are shown in figure 11. Using the global shape of the hand, the preliminary analysis showed that it was possible to recognize few of the chords played during the tests. The proposed prototype, consequently, considered each finger individually and solved the strings and frets components. During preliminary tests, the prototype was able to correctly recognize all fret positions. Due to the chosen camera view, the space between the strings is smaller for the high strings (E, B, G) than for the low strings (D, A, E), therefore affecting the recognition system accuracy. The low strings are recognized in almost all cases. For high string positions, the recognition rate is relatively lower.

8. CONCLUSIONS

Different strategies exist to retrieve the fingering information but actually none is able to provide a solution that respects the musician intent and naturalness with a sufficient degree of accuracy and precision. This article discusses new strategies to capture fingering of guitarists in real-time using low-cost video cameras. A prototype was developed to identify chords and fingering, based on finger-tracking and frets and strings detection. Results of the prototype are encouraging and open possibilities of studies on many aspects of the guitarist instrumental gesture orbiting the fingering problem, namely gesture segmentation, anticipatory movements, and bimanual synchronization. Applications of this research include automatic chord transcription, music education, automatic music generation and physical modeling.

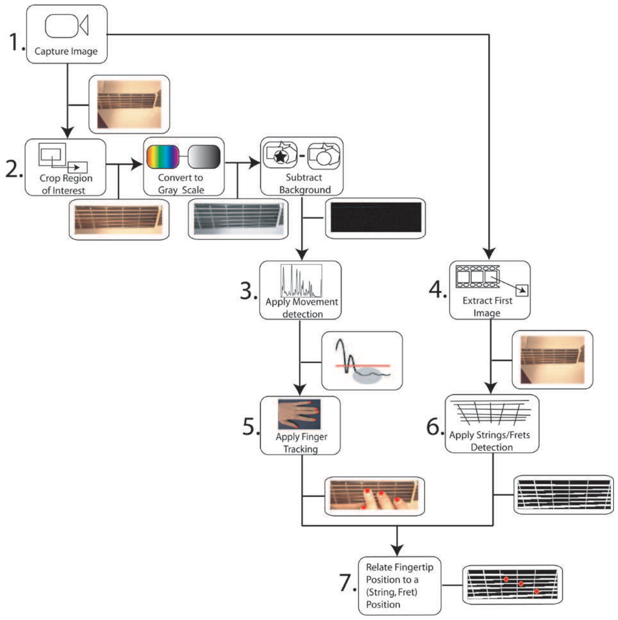


Figure 11. Second prototype algorithm

9. ACKNOWLEDGMENTS

The study on finger-tracking was realized at InfoMus laboratory, D.I.S.T., Universitá degli studi di Genova and has been partially supported by funding from the Québec Government (PBCSE), the Italian Ministry of Foreign Affairs, and by the EU 6 FP IST ENACTIVE Network of Excellence. The authors would like to thank all students and employees of InfoMus lab who ”lent a hand” for the realization of the tests. Special thanks go to Barbara Mazzarino, Ginevra Castellano for her help in compiling the results, Gualtiero Volpe for his contribution to the development of the EyesWeb blocks. Anne-Marie Burns would also like to thank Antonio Camurri for welcoming her as an internship student researcher.

The authors would like to thank the guitarists that participated in the tests: Jean-Marc Juneau, Riccardo Casazza, and Bertrand Scherrer. Thanks also to Mr. Donald Pavlasek, laboratory superintendent of the Electrical and Computer Engineering Department of McGill University for the conception and creation of the camera guitar mount and to Vincent Verfaille for his precious help for the latex formatting of this document.

10. REFERENCES

[1] A. P. Baader, O. Kazennikov, and M. Wiesendanger. Coordination of bowing and fingering in violin playing. *Cognitive Brain Research*, 23:436–443, 2005.

[2] A.-M. Burns and B. Mazzarino. Finger tracking methods using eyesweb. In *S. Gibet, N. Courty and J.-F. Kamp (Eds)., Gesture Workshop 2005 Proceedings, LNAI 3881*, pages 156–167, 2006.

[3] J. A. Canny. Computational approach to edge detection. *, IEEE Transaction on Pattern Analysis and Machine Intelligence*, 8(6):679–698, 1986.

[4] R. O. Duda and P. E. Hart. Use of the hough transformation to detect lines and curves in pictures. *Communications of the ACM*, 15(1):11–15, January 1972.

[5] K. C. Engel, M. Flanders, and J. F. Soechting. Anitcipatory and sequential motor control in piano playing. *experimental Brain Research*, 113:189–199, 1997.

[6] A. Gilardino. Il problema della diteggiatura nelle musiche per chitarra. *Il "Fronimo"*, 10:5–12, 1975.

[7] A. Gilardino. Il problema della diteggiatura nelle musiche per chitarra. *Il "Fronimo"*, 13:11–14, 1975.

[8] K. Hemmi. On the detecting method of fingertip positions using the circular hough transform. In *Proceeding of the 5th Asia-Pacific Conference on Control and Measurement*, 2002.

[9] P. V. C. Hough. Method and means for recognizing complex patterns. *U.S. Patent*, 3,069,654, 1962.

[10] H. Koike, Y. Sato, and Y. Kobayashi. Integrating paper and digital information on enhanceddesk: A method for realtime finger tracking on an augmented desk system. *ACM Transaction on Computer-Human Interaction*, 8(4):307–322, 2001.

[11] J. Letessier and F. Brard. Visual tracking of bare fingers for interactive surfaces. *Seventeenth An- nual ACM Symposium on User Interface Software and Technology*, 6(2):119–122, 2004.

[12] S. Paschalakis and P. Lee. Pattern recognition in grey level images using moment based invariant features image processing and its applications. *IEE Conference Publication*, 465:245–249, 1999.

[13] V. I. Pavlovic, R. Sharma, and T. S. Huang. Visual interpretation of hand gestures for human-computer interaction: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):677–695, 1997.

[14] D. Radicioni, L. Anselma, and V. Lombardo. A segmentation-based prototype to compute string instruments fingering. In *Proceedings of the Conference on Interdisciplinary Musicology*, Graz, 2004.

[15] C. Traube. *An Interdisciplinary Study of the Timbre of the Classical Guitar*. PhD thesis, McGill University, 2004.

[16] C. Traube and J. O. Smith III. Estimating the plucking point on a guitar string. In *Proceedings of the COST G-6 Conference on Digital Audio Effects*, Verona, Italy, 2000.

[17] J. A. Verner. Midi guitar synthesis yesterday, today and tomorrow. an overview of the whole fingerpicking thing. *Recording Magazine*, 8(9):52–57, 1995.