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 recordings of a guitarist. These first analyses led to some important findings about the design methodology of such a system, namely the focus on the effective gesture, the consideration of the action of each individual finger, and 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 on movement segmentation. Finally, these concepts were integrated into a prototype and a system for left hand fingering detection was developed.

1. INTRODUCTION

Fingering is an especially important aspect of guitar playing, as it is a fretted instrument where many combinations of string, fret, and finger position 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 noted [6] [7], specific fingering information is rarely indicated in musical scores.

Fingering can be deduced at several points 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 through score analysis. 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 with 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 hardly deal with situations where the musical intention is more important then the biomechanical optimum fingering.

Other systems retrieve the fingering during or after a human plays the piece. One of these approaches uses a Midi guitar. Theoretically, using a Midi guitar with a separate Midi channel assigned to each string, it is possible to know in real-time what pitch is played on which string, thus determining fret position. In practice however, Midi guitar users report several problems, including a variation in the recognition time from one string to another and the necessity to adapt their playing technique to avoid glitches or false note triggers [17].

An approach 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, finding the plucking point and then determining the string length to evaluate the fingering point. Shortcomings of this method are that 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 (and therefore the fingering evaluation) is better than one centimeter in the case of open strings but can be as high as eight 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 computational approach is inspired by the research on table-top applications [10][11] and by the research on three-dimensional finger-tracking [8]. These applications often use specific and expensive hardware (an 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 enables the tracking to each fingertip.

The first part of this article is a discussion about the preliminary manual and automated analyses of multiple-views recordings of a guitarist playing a variety of musical excerpts. The theoretical foundation of the Hough transform algorithm used is explained in the following section. Later, three studies are presented on the important aspects of visual analysis of guitar fingering, namely finger-tracking, string and fret detection, and movement segmentation. Finally a system integrating these three components is presented.

2. PRELIMINARY ANALYSIS

During the preliminary analyses, 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 maximum amount of information possible with a high 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, thus loosing complementary information. These analyses 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 can be seen in figure 1(a), the global view of the hand is ideal for its richness in gestural information. This view permits observation of 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 content since the player's 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, information about the right hand, as well as postural and facial gesture, are lost. On the other hand, 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 assist 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 ability to detect strings and frets, and the potential occurrence of the finger occlusion problem. Moreover, this view shows the fingers proximity to the string; it may therefore be possible to know whether the string is pressed or not by the guitarist. Another potential advantage of this view is that it is close to the musician's view of the neck when playing. This does not specifically pertain to finger-tracking but it may perhaps be useful in a system designed for educational purposes.

(a) Global view with zooms

(b) Front view of the left hand

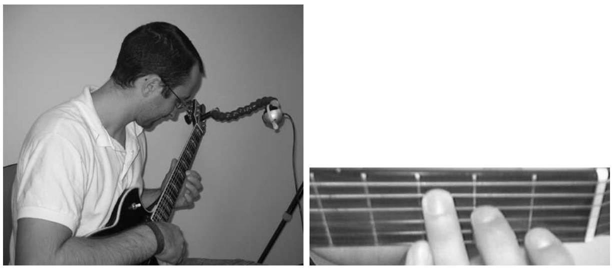
(c) Top view of the left hand

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

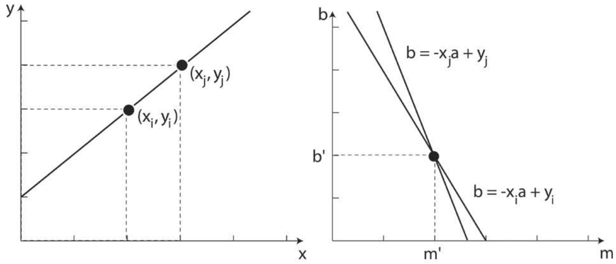
The top view (figure 1(c)) was retained for its interesting characteristics with respect to the problem, namely a detailed view of the fingers, the possibility for string and fret detection, and the ability to observe finger-string proximity. However, slow motion observations of the video recording showed that the neck is subject to ancillary movements of the performer. Notably, rotational movements in three dimentions were 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 variations due to translational movements. Preliminary automated tests have shown that this type of movement can influence the computer's capacity to correctly identify fingering. Consequently, the tripod was replaced 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 musical excerpts against top view images of previously recorded chords played by the same performer. In the computer algorithm, the hand was segmented from the rest of the image, first by the manual selection of the region of interest, a rectangular region around the hand, then by applying the Canny edge detection algorithm [3] on the threshold image of that region. Time segmentation was also performed in order to evaluate only stable images of chords. In other words, the



(a) Camera mount (b) Camera view

Figure 2. The guitar mount that was used to eliminate the ancillary gesture problem: (a) The camera mount installed on an electric guitar. (b) The camera view on a classical guitar. In this example, the camera is placed 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.

transitory phases of preparation and retraction were eliminated based on the assumption that they are high motion phases and only the nucleus phase was retained and evaluated. This was done by establishing a fix threshold on the motion curve generated by the computation of the difference of pixel between each images of the sequence. The images obtained after that process were converted to vectors by applying the Hu moments algorithm [12]. Eleven chords of the first recording were chosen and stored in a knowledge base in the form of Hu moments vectors. A chord was recognized when a chord of the knowledge base matched it with a sufficiently high proximity factor. These preliminary tests permitted to identify three main issues of the guitarist fingering problem:

1. Using an appearance-based 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.

At its conclusion, the preliminary analysis provided insight about how to visually acquire guitar fingering information. 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 examined the use of the linear Hough transform for string and fret detection and a third one explored 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 approximated by geometric shapes. In this paper, two kind of Hough transform are used:

1. The linear Hough transform is used to detect the guitars strings and frets;
2. The circular Hough transform is used to detect fingertips, whose ends can be approximated with a semicircular 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:

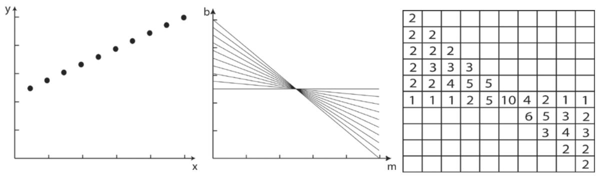
$$y_i = mx_i + b \tag{1}$$

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:

$$b = -x_i a + y_i \tag{2}$$

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 parameter space therefore needs to be discretized. This is done by dividing the parameter space in accumulator cells. Each point is then tested for all possible m values in 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 illustrates an example of ten points that can be linked into a line. Figure 4(b) displays the solutions for the line in 4(a). It can be observed



(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.

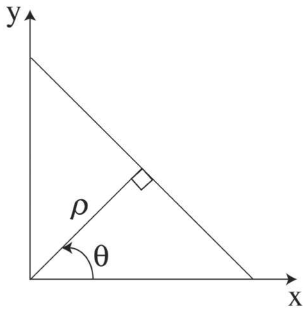


Figure 5. Normal parametrization of a line given by equation 3 where θ and ρ are fixed parameters

that collinear points in the xy space correspond to 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 4(c).

One problem with this approach is that the intercept and the 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].

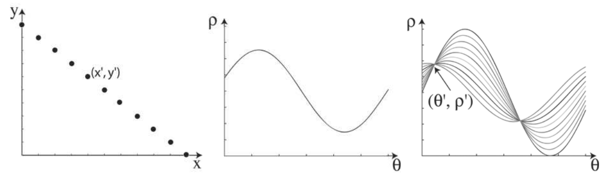
As illustrated in figure 5, a line can be described by the angle θ of its normal and by its distance ρ from the origin using the following equation:

$$x \cos\theta + y \sin\theta = \rho \tag{3}$$

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

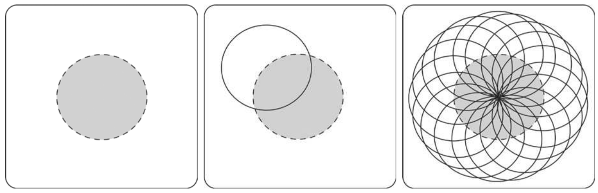
3.2. The circular Hough Transform

The concept underlying the circular Hough transform introduced by Duda and Hart in 1972 [4] is quite similar to the linear one. As described in that article, the Hough transform can be applied to 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.



(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.

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

$$(x - a)^2 + (y - b)^2 = r^2 \tag{4}$$

When possible, it is advantageous to reduce the parameter 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.

$$x = a + r \sin\theta \tag{5}$$
$$y = b + r \cos\theta \tag{6}$$

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.

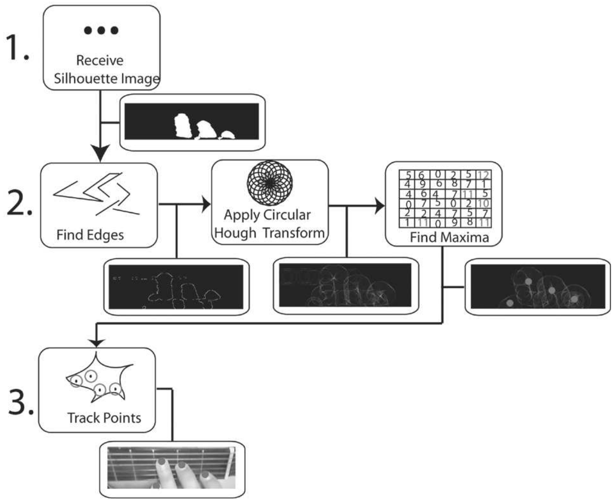


Figure 8. Fingertips detection using the circular Hough transform algorithm

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]. Among all the tested algorithms, the circular Hough transform was retained for the guitarist fingering problem due to several interesting characteristics:

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. An edge image is obtained by applying the Canny edge detection algorithm [3] on the silhouette images (figure 9(a)). The circular Hough transform algorithm makes use of 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 the center of the fingers' ends (figure 9(b)).

5. STRING AND FRET DETECTION

By tracking the fingertips it is possible to know the location of each finger is in space. In the case of fingering on the guitar, this space can be defined in terms of string and fret coordinates. The detection of strings and frets in the image is consequently a crucial step. Figure 10 shows the string and fret detection algorithm. Prior to

(a) Edge image (b) Circular Hough image

Figure 9. Detection of maxima at the fingertips: (a) Canny edge-image of the hand; (b) Circular Hough image.

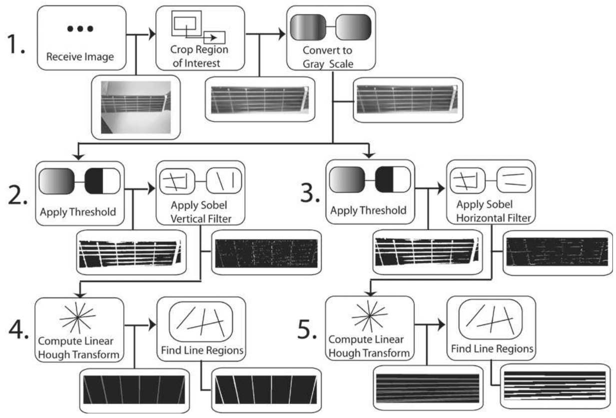
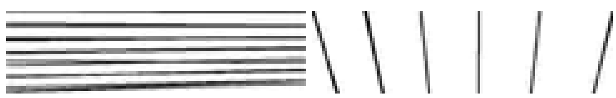


Figure 10. String and Fret detection using the linear Hough transform algorithm

detection stage, the region of interest (in this case the guitar neck) must be located in the image. Once the neck has been located, the strings and frets are segmented from the grayscale neck image by applying a threshold. A vertical and a horizontal Sobel filter are applied on the threshold image to accentuate the vertical and horizontal gradients. A Linear Hough Transform is then computed on the two Sobel images. The linear Hough transform allows detection of linearity in group of pixels, creating lines (figure 11). These lines are then grouped by proximity in order to determine the position of the six strings and of the frets. Once this is done, it is possible to create a grid of coordinates to which fingertip positions can be matched.

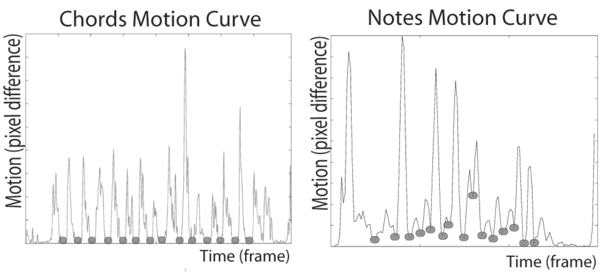
6. MOVEMENT SEGMENTATION

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 significant finger positions (see figure 13 line 3). Movement segmentation is



(a) Strings (b) Frets

Figure 11. Linearity detected in the string and fret regions: (a) String regions; (b) Fret regions.



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

Figure 12. (a) Motion curve of a guitarist playing chords (b) Motion curve of a guitarist playing notes

used to separate the nucleus phase of the gesture from the preparation and retraction phases. 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 analyses, movement segmentation was done by applying a threshold on the motion curve (figure 12 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. However, in other playing situations, such as when playing a series of notes, the separation between the movement transitory phases and the nucleus is not that clear (figure 12 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 occur during the nucleus of the action-finger. Thus the motion is not serial, and consequently the global motion curve does not exhibit clear global minima as in the case of simple chord fingerings. 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. Local minima are found by computing the second derivative of the motion curve. As the prototypes work in real-time with discrete data, this is done by subtracting the signal from its delayed version twice.

7. PROTOTYPE

The prototype was designed to fulfill the requirements for a fingering recognition system highlighted by the preliminary analyses. The focus on effective gestures is partially realized at hardware level by affixing the camera to the guitar neck, thereby eliminating the motion of the neck caused by ancillary gesture. 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 in pixel values between each frame. The action of each individual finger is considered using the finger-tracking algorithm described above. Gesture is recognized by matching the fingertip positions to the string and fret grid of coordinates, thus not relying on any knowledge base. The details of the algorithm are shown in figure 13.

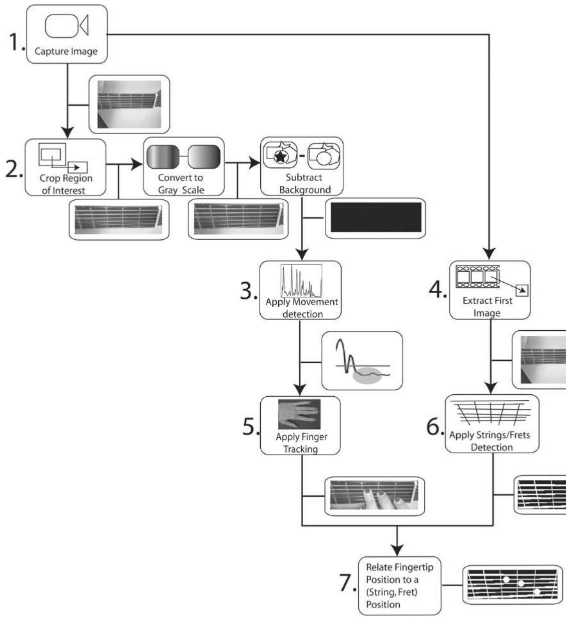


Figure 13. Second prototype algorithm

During preliminary tests, the prototype was able to correctly recognize all fret positions. Due to the placement od the camera, the space between the strings is smaller for the high strings (E, B, G) than for the low strings (D, A, E), affecting the accuracy of the recognition system. As demonstrated in [2], the circular Hough transform has an accuracy of 5 +/- 2 pixels with respect to the color marker references. The resolution of the camera used in this prototype is 640x480 pixels, therefore giving a 610x170 pixels neck region. The distance in pixels between the first and second string is of 12 pixels at the first fret and 17 at the fifth fret. Between the fifth and sixth strings, the distance in pixels is 16 and 20 pixels for the first and fifth fret, respectively.

Since the chosen algorithm attributes the string position to the finger by proximity, in the worst case the finger-tracking algorithm error exceeds half the space between the higher strings, therefore confusion happens. However, since this problem does not happen with lower strings were the distance between two strings is greater, the problem could be solved with an higher resolution camera. Another limitation is that in the current system only the first 5 frets are evaluated, but this could be solved with a wide angle camera. One problem that cannot be easily solved by changing the hardware is finger self occlusion. This problem only rarely happens, but exists in the case of fingerings were two fingers play at the same fret, for example in the case of C7 and Dm7. In future developments, this problem could potentially be solved by estimating the fingertip position using the finger angle.

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 series of notes based on finger-tracking and fret and string detection. It recognizes fingerings by matching fingertip positions to the string and fret grid of coordinates, thus not relying on any knowledge base. Results of the prototype are encouraging and open possibilities of studies on many aspects of the guitarist instrumental gesture, namely gesture segmentation, anticipatory movements, and bimanual synchronization. Applications of this research include automatic chord transcription, music education, automatic music generation and physical modeling.

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 first author 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. hanks also to Mr. Donald Pavlasek, Electrical and Computer Engineering, McGill University, for the conception and implementation of the camera guitar mount.

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.