Vol.2011-EC-19 No.5
2011/3/26

# 紙楽譜を用いた演奏メディア onNote のためのマーカレス楽譜認識の提案

山本祐介† 内山英昭†† 筧康明†††

本稿では，紙楽譜を用いた演奏メディア onNote のためのマーカレス楽譜認識の評価実験の結果を報告する．我々は，カメラに対して紙の楽譜をかざし，動かすことで，その楽譜の種類や位置などに応じた音楽を演奏できるインタラクティブシステム onNote を開発している．システムの実装に向けて LLAH(Local Likely Arrangement Hashing)を応用し，マーカ等を添付しない紙楽譜に対して，種類や姿勢をリアルタイムに検出するための画像認識法を提案し，楽譜上でポインティングされた点と音との対応付けを可能にした．本稿では，システムの概要を述べると共に，この画像処理手法に関する評価実験として，楽譜の識別性，処理速度，カメラ位置のロバスト性に関する実験を行った．

# A Proposal on a Markerless Recognition of Physical Scores for onNote

Yusuke Yamamoto† Hideki Uchiyama† and Yasuaki Kakehi†††

In this research, we have developed a novel musical interface system named ``onNote.''.In this system, physical markerless musical scores are used as instruments to play music intuitively. Users can enjoy playing music in various ways: moving, rotating pointing the scores. In this research, we developed the method for extracting the stable keypoints on the score and combined it with LLAH for key-point matching then the system identifies the score by referencing the score database. In addition, by using these methods, the system recognizes the position and rotation of the score and let us control the sound pointed in real-time. In this paper, we describe the outline of the system and report the result of examination: discrimination of the scores, processing speed and robustness to the camera position.

## 1. はじめに

レコードや CD に代表される演奏メディアは，ただ内包している音楽データを再生するのみならず，DJ が行うスクラッチなどの簡易な操作を加えることで，多様なエフェクトや表現につなげることができる．これは，既存のメディアに限らず，近年ではデジタルメディアを用いて様々な演奏メディアが提案され，音楽を作る行為は特別な教育を受けた人のみならず，より多くの人が楽しめるものとなってきた．しかし，レコード盤だけを見ても，その中のどこにどのような音が入っているのかが分からないように，これまでの演奏メディアの多くは記録された情報としての音楽と，それを記録する媒体となる物理的素材の間に必ずしも明確な対応づけがなされてこなかった．このような背景のもと，より直観的な演奏メディアの実現に向けて我々は，音が記号として記録された物理的媒体として楽譜（五線記譜法を用いた楽譜）に着目した．紙の楽譜を直接インタフェースとし，その楽譜の種類や位置姿勢等に応じた音楽を演奏できるインタラクティブシステム onNote を開発している[1]. 図 1 のようにユーザは，机の上に置かれた照明に内蔵されたカメラとプロジェクタの前に楽譜をかざし，動かすことで，自分の好きな速度で，好きな位置の音楽を演奏できる．楽譜は視覚的に音

図 1: onNote
Figure 1: onNote.

† 慶應義塾大学大学院政策・メディア研究科
Graduate School of Media and Governance, Keio University
†† フランス国立情報学自動制御研究所
INRIA Rennes
††† 慶應義塾大学環境情報学部
Faculty of Environment and information Studies, Keio University

楽の存在や構造を読み取ることができる．また，楽譜が印刷された紙は私達の身近な素材であり，折る，曲げる，重ねるといった多様な使い方ができる．これらの操作を演奏手法として積極的に取り入れることで，高い直感性を有する演奏メディアの実現を目指している．

onNote では，このような自由な演奏方法を可能にするために，マーカレスの楽譜に対して，カメラ画像からの自然特徴点のマッチングを用いた楽譜画像検索を行う．本稿では，以下に楽譜認識に関する関連研究をまとめた後に，システムの概要，この画像処理手法に関する性能評価の結果を報告する．

## 2. 関連研究

紙楽譜を認識する方法は，従来からいくつかの取り組みがなされている[2]．それらの多くは，楽譜に含まれる演奏情報をデジタルで扱えるデータに変換することを目的としている．例えば，楽譜を MIDI 情報に変換するための技術である．変換の際のアルゴリズムとしては，音符や記号を抽出し，位置や種類の認識を行うものである．現在は，精度と変換速度が向上し，市販のソフト[3]でも，１枚の楽譜を数秒でデジタル情報に変換できる．しかし，カメラ画像からの認識や紙が変形し隠れが生じた場合の認識などには不向きであり，今回のように楽譜をカメラの前で様々に動かして楽譜の識別とカメラの位置姿勢推定をリアルタイムに行うような目的には適さない.そこで，本研究では楽譜の中に記された記号情報を用いて自然特徴点検索のアプローチで楽譜の姿勢や ID をリアルタイムに検出する手法を提案・実装している．

楽譜に限らず，マーカレスで物体の状態認識と特定を行う手段として，自然特徴点を用いた手法が挙げられる．自然特徴を用いた特徴点マッチングとは，実世界に既に存在する特徴を用いたパターン認識の方法であり，ARToolkit [4]のように人為的に画像処理のためのマーカを貼付する必要がない．また，オクルージョンに強く，幾何学的に複雑な変形も認識できるという利点もある．自然特徴点のマッチングとして代表的なものとして，SIFT [5]や SURF [6] があが，これらの手法で用いられる特徴点，特徴量の計算はリッチなテクスチャに適した方法であり，楽譜のような二値画像に対して不向きである．一方，二値画像の検索に適した手法としては，点や線などの幾何学位置関係を利用した GH(Geometric Hashing) [7] や, 文章の検索法として用いられる Local Likely Arrangement Hashing (LLAH) [8] がある．この中で，GH は計算コストが非常に大きく，リアルタイムでの情報提示は難しい．一方，LLAH は文章に特化した特徴点抽出のために設計されており，検索処理が高速であるとともに，射影変形や照明変化に対してもロバストである．さらに文献 [9] では，LLAH の検索アルゴリズムの高速化，特徴点のトラッキング手法が提案され，処理速度とカメラの位置に対する自由度が改善された．このような背景から，本研究では LLAH をベースに楽譜に適した特徴点抽出を行う楽譜画像検索を用いて紙楽譜の認識を行っている．

## 3. onNote における楽譜画像検索

### 3.1 onNote 概要

onNote では, ユーザは紙の楽譜を直接演奏に用い，楽譜の音を紙の特性を用いて操作できる．例えば，紙の楽譜をカメラの前にかざし，動かすことにより，楽譜上の任意の場所の音楽を任意の速度で演奏するなどの演奏を行うことができる．このような機能を実現するために，図 2 のような処理を行う．

まず，システムでは，カメラから入力された楽譜画像に対して楽譜特徴点抽出を行い，特徴点に変換する．特徴点に変換された画像は，LLAH による幾何学特徴量計算を用いた特徴量の記述，さらにはデータベースとのマッチングが行われる(Keypoint Matching by LLAH)．マッチングの結果，楽譜の ID とカメラに対する位置姿勢が計算される(Camera Pose Estimation)．予め登録された楽譜の座標に MIDI 情報を対応づけておくことで，これらの情報からそれぞれの楽譜の位置姿勢を用いた演奏が可能になる(MIDI Selection)．以下，それぞれの処理に関して詳しく述べる．

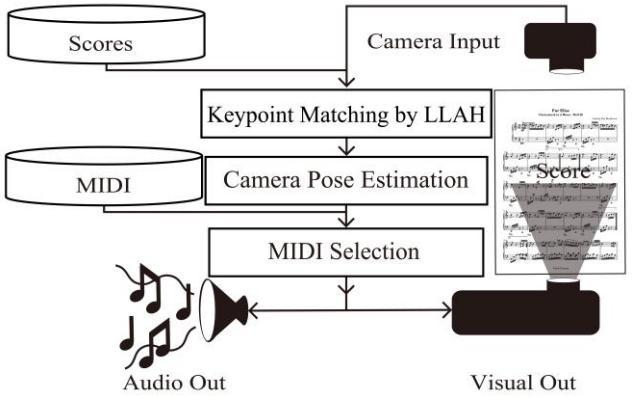


図 2: 処理の流れ
Figure 2: Algorithm Overview.

### 3.2 楽譜特徴点抽出処理

LLAH は，一般的には文章画像検索に用いられる．本システムでは，この手法を楽譜検索に応用するために，画像中の楽譜記号を特徴点として用いる．一般的に楽譜では，音符等の記号は五線をまたいでいるため，通常の二値化処理を施しただけでは，五線によって各記号が連結された状態で抽出されるため，１つの楽譜から数個の特徴

点しか得られないことになる．これに対し，自然特徴点検索では特徴点を用いて検索を行うため，１つの画像ができる限り多くの特徴点を持っていたほうが望ましい．

この問題を解決するために，本システムでは，楽譜の五線や音符の符幹などを含む細い線を削除し，音符記号に含まれる符頭や符尾などある程度の大きさ有する部分を特徴点とするアルゴリズムを導入した．具体的には，図 3 のように入力画像をグレースケールに変換，ガウシアンフィルタでぼかし，適性 2 値化で画像の濃く残った部分だけを抽出し，その重心を特徴点とする処理を施す．提案システムでは，必要な情報をプロジェクタで直接楽譜に投影する．そのため，プロジェクタでの投影映像を単色とし，楽譜特徴点抽出処理の入力画像をグレースケールに変換する際に，その色を除き，認識の影響を避ける．

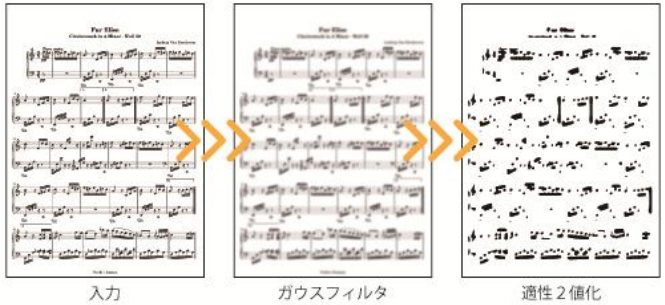


図 3 楽譜特徴点抽出法
Figure 3: Extracting the keypoint on the score.

### 3.3 LLAH

LLAH では，各特徴点に対してその近傍の点の幾何学位置関係を用いた特徴量によって特徴点検索が行われる．ある特徴点 p 点に対して，n 点の近傍点を取得し，近傍 n 点中の m(m≦n, m≧4)点を選ぶ．nCm 個の組み合わせそれぞれに対して mC4 個の点の組み合わせを求め，4 点から計算されるアフィン不変量を離散化レベル数 k に基づいて離散化した値の列(r(0),r(1),...,r(mC4-1))を特徴量として用いる．この特徴量をもとにそれぞれの特徴点がハッシュサイズ Hsize のハッシュ表に登録されている．ハッシュ表のインデックスは以下に示すハッシュ関数で計算される．

$$Hindex = \left( \sum_{i=0}^{mC4-1} r(i)\, k^i \right) Hsize \qquad (1)$$

実際にハッシュ表には，文章の識別番号，点の識別番号，点の座標，不変量の離散値化した特徴量列(r(0),r(1),...,r(mC4-1))が登録される．検索は検索画像中の特徴点に対して行われ，特徴点のメタ情報から楽譜の ID や位置姿勢が計算される．

演奏の際には，LLAH の処理後計算された ID と位置姿勢から，事前に用意した MIDI のデータベースに照合することで楽譜の指定した場所に対応する MIDI が再生される．このデータベースは，MIDI メッセージを楽譜画像の小節ごとの始めの音の鳴り始めと終わりの音の鳴り終わりを区切りとして楽譜の xy 座標に割り当てて作成する．

## 4. 実験

### 4.1 実験概要

上記のような手法で設計・実装した onNote システムに関して，下記の 3 つの観点からその楽譜画像検索の性能に関する評価実験を行った．

一つ目は，識別性に関する実験である．本手法を用いることで楽譜の種類を見分けられるのかについて，実験を行う．二つ目は，処理速度に関する実験である．本システムはリアルタイムでの演奏を想定しており，その動作速度は重要なパラメータとなる．三つ目は，楽譜に対するカメラ位置のロバスト性に関する実験である．拡大／縮小，傾きに関して位置関係を変えながら，システムの振る舞いを調べる．

まず，実験に用いる楽譜として 101 枚の印刷楽譜を用意した．これは，楽譜公開サービス free-scores.com（http://www.free-scores.com/）において，楽器をピアノと指

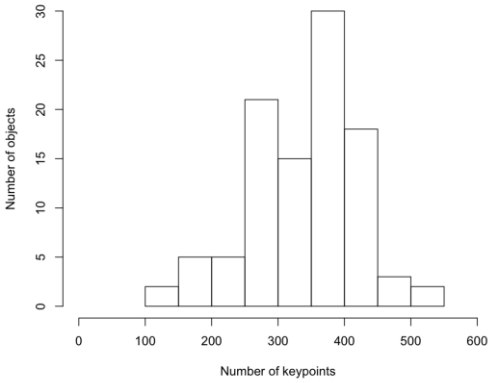


図 4：選定楽譜中の特徴点数に応じた楽譜数分布
Figure4: A histogram of scores according to the number of keypoints.

定して検索した Most download sheet music のうちから 16 個を選び, その中から 101 枚を任意に選定した．選定した 101 枚の楽譜は，含まれる特徴点の数の分布で表すと図 4 のようになった．

LLAH の処理はパラメータによる影響が大きい．上記のように LLAH には，n，m，k，Hsize という 4 つのパラメータがある．それぞれのパラメータがどのように処理に影響するのかということに関しては文献[10]を参照されたい．本投稿では，101 枚の楽譜に対して実験を行ったため，Hsize=$2^{32}$として固定し，識別性，処理速度では，下記の 6 通りのパラメータを試した．(n,m,k)=(6,5,4), (7,5,4), (7,6,4), (6,5,3), (6,5,4), (6,5,5) ．ロバスト性の実験に関しては，実験(n,m,k)=(7,5,4)で検索できたもの使って実験を行った．

これらの実験は全て，PC（MacBook，2.4GHz intel Core 2 Duo，2GB(1GB SO-DIMM×2) 1,066MHz DDR3 SDRAM）とカメラ（UCAM-DLU130H）を用いて，室内通常照明条件化のもとで行った．カメラの入力解像度は，640 [pixels]×480 [pixels]．登録楽譜は，楽譜全体がちょうどカメラフレームに収まるように，楽譜に対してカメラを 40cm 離し，水平に固定し撮影した．その際，楽譜の登録画像の解像度は 640 [pixels]×452 [pixels]とした．

### 4.2 実験 1: 識別性に関する実験

101 枚の楽譜をデータベースに登録し，40cm にカメラを固定し，LLAH の各パラメータセットで楽譜の識別の可否を調べた．この実験の結果は表 1 のようになった．

表 1　識別性と処理速度の結果
Table 1: The result of experiment about discrimination and processing time.

| n | m | k | 精度[枚] | | | | 処理速度[msec/frame] | |
|---|---|---|---|---|---|---|---|---|
| | | | 検索可 | 不安定 | 複数枚 | 検出不可 | 特徴点抽出 | 特徴点検索 |
| 7 | 6 | 4 | 64 | 8 | 28 | 1 | 20 | 22.5 |
| 7 | 5 | 4 | 66 | 7 | 28 | 0 | 24 | 191.5 |
| 6 | 5 | 4 | 71 | 10 | 17 | 3 | 29 | 42 |
| 6 | 5 | 3 | 67 | 11 | 21 | 2 | 29 | 168.5 |
| 6 | 5 | 5 | 67 | 7 | 21 | 6 | 25.5 | 16.5 |

認識が不安定とは，楽譜を動かさない状態では認識されないが，左右上下にわずかに楽譜を動かすことで反応したり，何フレームかに一度認識できたケースを指す．複数枚認識されるとは，図 8 のように一枚の楽譜に対して，複数枚の楽譜検索結果得られたケースを指す．

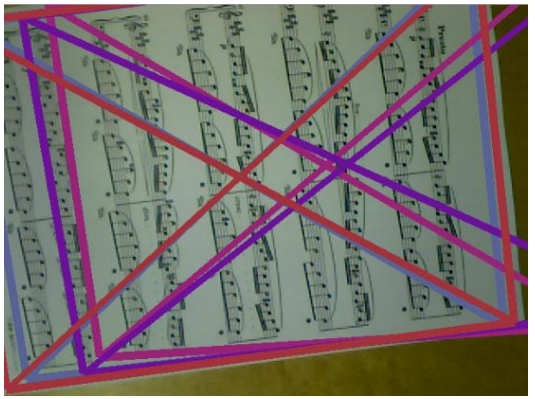

図 5: 複数枚検索された例
Figure 4: An example of capture result retreiving multiple scores.

検出不可とは，まったく楽譜が認識できず，間違った検出結果も表示されなかったということである．パラメータセットにも依存するが，101 枚中 0〜6 枚であった．

### 4.3 実験 2：処理速度に関する実験

101 枚の中から 1 枚の楽譜を選び，カメラを楽譜から 40cm の位置に正対して固定し，LLAH の各パラメータセット（n, m, k）のもとで特徴点を検出した際の画像検索に要する時間を 100 フレームに対して測定し，1 フレームにかかる計算時間の平均を算出した．この実験の結果も表 1 に示す．

### 4.4 実験 3：カメラ位置のロバスト性

LLAH のパラメータセットを(n,m,k)=(7,5,4)に設定し，楽譜は実験 1 の際において

表 2　拡大縮小に応じた識別精度
Table 2: The result of discrimination according to the distance between the camera and the score.

| 高さ[cm] | 精度[枚] | | |
|---|---|---|---|
| | 検出可 | 不安定 | 検出不可 |
| 25 | 24 | 12 | 30 |
| 30 | 61 | 5 | 0 |
| 35 | 66 | 0 | 0 |
| 45 | 66 | 0 | 0 |
| 50 | 6 | 7 | 53 |

このパラメータセットで検出できた 66 枚を対象に行った．

まず，スケールに関するロバスト性を調べるため，楽譜から 25cm，30cm，45cm および 50cm の位置にカメラを水平に固定し，上記の 66 枚の楽譜を用い，その中の何枚を検出できるかを測定した．図 6 はカメラと楽譜の距離がそれぞれ 25cm，50cm の時のカメラ画像である．

その結果，拡大縮小に関しては表 2 のような結果となった．上述の通り今回の実験ではカメラと楽譜距離が 40cm の時に最も検出されやすいようにデータベースの解像度などが合わせてあるため，実験結果より，縮小よりも拡大に対して認識がうまく行われることが分かった．

次に，楽譜の中心からカメラレンズまでの距離を 40cm に固定し，楽譜の中央を中心として，45 度，60 度，75 度とカメラを傾けた際に，同様に 66 枚中何枚の楽譜を検出できるのかについて測定した. 図 7 は, それぞれの傾きの際のカメラ画像である．

結果を表 3 に示す．角度が増すと検出が不安定になる枚数が増してくるが，楽譜に対してカメラを 45 度程度傾けてもある程度の精度で認識が行われるということが分かった．

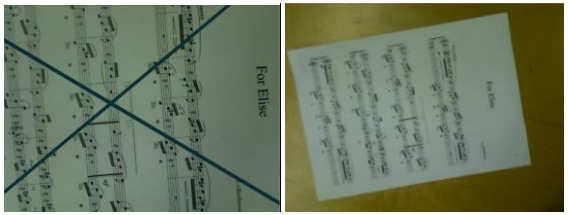


図 6: 距離を変化させた際の入力画像（左 25cm，右 50cm）
Figure 5: Appearance of the input image according to the distance between the camera and the score (left 25cm, right 50cm).

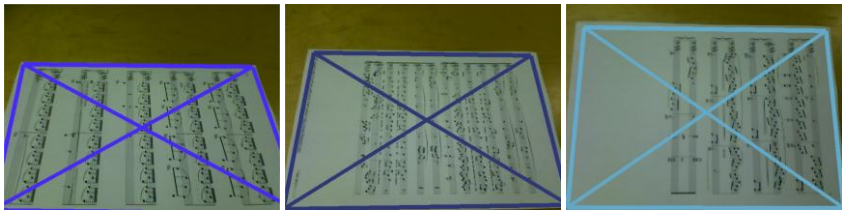

図 7: 傾きを変化させた際の入力画像（左 45 度，中 60 度，右 75cm）
Figure 6: Appearance of the input image according to the angle between the camera and the score (left 45 degrees, center 60 degrees, right 75 degrees).

表 3 傾きに応じた識別精度
Table 3: Table 2: The result of discrimination according to the angle between the camera and the score.

| 角度[度] | 精度[枚] | | |
|---|---|---|---|
| | 検索可 | 不安定 | 検出不可 |
| 75 | 63 | 1 | 2 |
| 60 | 59 | 5 | 2 |
| 45 | 55 | 8 | 3 |

### 4.5 実験考察

識別性の実験では，パラメータを n=7，m=5，k=4 としたとき 101 枚中 66 枚が検索できるなど，ある程度識別性のある楽譜認識方法であることが分かった．しかし，認識が不安定なもの，一枚の楽譜に対して複数枚の検索結果が出てくる楽譜があるという問題が分かった．認識が不安定なものに関しては，カメラに対して楽譜を上下させるなどの操作を行うと反応するため，カメラから入力された画像とデータベースとの画像がノイズやカメラの歪みの影響で異なるため，特徴点抽出の際にうまくデータベースと同じように特徴点抽出が行われなかったことが原因と考える．それぞれのカメラ，楽譜に合わせたカメラキャリブレーション，データベースへの登録作業が必要であるといえる．

一枚の楽譜に対して複数枚の検索結果が計算される楽譜に関しては，検索結果に必ず正しい検索結果が含まれており，同じ楽曲内の違うページの楽譜が検索されることから，楽譜同士が非常に似た記号配列を持っていることが原因と考えられる．図 8 から分かるように，左手のパートがほぼ同じ音符配列になっている．現在の LLAH の幾何学特徴量の記述では，このような問題が避けられない．この問題に関しては，射影変化のずれや，特徴量の記述方法を工夫する，既存の楽譜認識と組み合わせるなどの解決策が考えられる. 今回使用した楽譜では, Fantaisie Impromptu-Op.66(F. chopin) などの楽譜同士がこのような結果となった．

処理速度に関しては，全体的に特徴点検索の際に多くの時間を要していることがわかる．特徴点検索に要する時間は，PC のスペックに依存するところが大きく，今回使用した PC では, 音楽性の高い演奏をするためにはフレームレートが低いと考える．しかし，この部分は，PC スペックを上げることで解消されると考える．

実験 3 のロバスト性に関する実験のうち，拡大縮小に関しては，縮小よりも拡大に対して強度が高いということが分かった．実験では，25cm から 50cm の距離の範囲で測定したが，この範囲を外れた距離では，検索正解率が著しく低下した．この原因は，カメラ画像が拡大縮小し，カメラ画像から得られる楽譜画像の解像度が変化する

のに対し，データベースの画像の解像度は一定であり，特徴点抽出処理において特徴点が正しく得られなかったことが原因であると考える．この問題に足しては，データベース登録の際にピラミッド画像を用意しておき，カメラの解像度にあったデータベース画像で楽譜画像検索を行うなどの解決策が考えられる．一方の傾きに関する実験結果では，楽譜に対して楽譜を 45° 傾けた状態でも十分に検索可能であることが分かった．今後我々の先刻研究で提案したトラッキング手法などを取り入れることで，さらにカメラに対する自由度を上げることができると考える．また，楽譜をカメラに対して 45° 傾けたときにひとつの楽譜に対して間違った位置姿勢が認識されることが分かった．この問題に対しては，識別性の実験の際に複数枚の検索が得られた楽譜と原因は同じで，一枚の楽譜に似た記号配列がみられることが原因と考える．このような場合も，識別性の実験の際に複数枚の検索結果が得られた楽譜と同様の対策が必要である．また，今回の実験 1，実験 2 をまとめた表から分かるように，楽譜の識別性と，処理速度はトレードオフの関係にあるため，今後演奏方法に合わせたパラメータのチューニングが必要であると考える．

## 5. おわりに

本投稿では，紙楽譜を用いた演奏メディア onNote のシステム概要を述べ，提案した楽譜画像検索の評価実験を識別性，処理速度，カメラ位置のロバスト性の観点で行った．評価実験の結果から，データベースに登録した 101 枚の楽譜のうち，半数以上を正確に認識できることが分かった．また，カメラと楽譜の間の距離や角度の変化に対してもある程度のロバスト性を有していることが分かった．

一方で，楽譜では，同じ楽曲内，一つの楽譜内で似たような記号配列があるという楽譜特融の問題から，現在の LLAH を応用し，特徴点抽出部分を楽譜に特化させた現在のアルゴリズムでは，一枚の楽譜に対して複数の検索結果が得られたり，一枚の楽譜で違った位置姿勢が計算されてしまうという問題があることが分かった．処理速度の実験からは，今回はリアルタイム処理としては処理時間がかかっていたが，コンピュータの構成を見直すことで向上が見込まれる．

今後は，距離や角度の変化を活かした演奏方法など新たなインタラクション手法の提案に加え，楽譜画像検索の評価実験に対する改善を行っていく必要がある．

**謝辞**
本研究の一部は，独立行政法人 情報処理推進機構 未踏 IT 人材発掘・育成事業（2010 年度）の支援を受けて行った．

## 参考文献

1)山本祐介, 内山英昭, 篤康明:``onNote: カメラ画像による紙楽譜認識を用いた演奏メディア”, 情報処理学会，インタラクション 2011(2011).
2)加藤博一，井口征士：``小節単位処理に基づいたピアノ楽譜の自動認識”，電子情報通信学会論文誌，vol.J71-D, No.5，pp.894-901(1988).
3)株式会社河合楽器製作所: http://www.kawai.co.jp/cmusic/products/sm/，（2011 年 2 月現在）.
4)Kato, H. and Billinghurst, M.: ``Marker Tracking andHMD Calibration for a video-based AugmentedReality Conferencing System”, Proc IWAR’99, pp.85-94 (1999).
5)Low, D.G.: ``Distinctive image feature from scaleinvariant key points”, IJCV, vol.60, pp91-110, 2004.
6) Bay, H., Ess, A., Tuytelaars, T. and Gool, L.V.:``SURF: Speeded Up Robust Features”, CVIU, vol.110, pp.346-359 (2008).
7)Lamdan, Y. and Wolfson, H.J.: ``Geometric Hashing: A General And Efficient Model-based Recognition Scheme”, Proc. ICCV, pp238-249 (1988).
8)中居知弘，廣瀬浩一，岩村雅一：``処理速度とメモリ効率の改善された LLAH によるカメラベース文章画像検索法”，画像の認識・理解シンポジウム論文集，pp. 1252- 1259 (2008).
9)Uchiyama, H., Saito, H., Servieres, M. and Moreau, M.: “AR GIS on a physical map based on map image retrieval using LLAH tracking”, Proc. MVA, pp382-385 (2009).
10)中居友弘，黄瀬浩一，岩村雅一：``特徴点の局所的配置に基づくディジタルカメラを用いた高速文章画像検索”，電子情報通信学会論文誌，Vol. J89-D，No. 9，pp. 2045-2054 (2006).