

Title	ハイブリッド処理を用いたフィンガープリント情報の 高速検索に関する研究
Author(s)	熊坂, 史彬
Citation	
Issue Date	2008-03
Type	Thesis or Dissertation
Text version	author
URL	<a href="http://hdl.handle.net/10119/4318">http://hdl.handle.net/10119/4318</a>
Rights	
Description	Supervisor:井口寧, 情報科学研究科, 修士

# ハイブリッド処理を用いた フィンガープリント情報の 高速検索に関する研究

熊坂 史彬 (0510033)

北陸先端科学技術大学院大学 情報科学研究科

2008 年 2 月 7 日

キーワード: ハイブリッド, 並列, ハードウェア, FPGA, 検索.

## 1 あらまし

近年における高速ネットワークの急速な普及によって、マルチメディアコンテンツをデジタルデータとして容易に扱えるようになった。これにより著作権に対する認識も従来と比較し重要視されるようになった。デジタルデータの内容から、一意のコンテンツデータを特定することが可能な技術などを総合し、デジタル著作権管理技術と呼ぶ。このうちの一つにフィンガープリント技術がある。これはコンテンツデータの信号自体に特徴量を見出し、これをコンテンツIDとする技術である。検出したIDは既存のデータベースを検索し、識別に至る。フィンガープリントIDを検出する過程はハードウェアによって高速に処理できることが明らかになっている。磯永らによれば、検出する過程をハードウェア化することで314msでの処理が可能になっている。ハードウェアのみによってフィンガープリント技術を実現しようとする場合、検索の過程で検索曲数の少なさが問題となる。またこれとは別にソフトウェアのみによってこれを実現しようとする場合、ソフトウェアの検索時間が問題となる。本研究では、特にこのフィンガープリント技術に着目し、コンテンツの流通時に著作権管理を行う手法を考える。

## 2 提案手法

提案手法としてはソフトウェアとハードウェアのハイブリッド処理を利用する事によって、互いの不足する部分を補う事を狙った。短い検索時間を維持しつつ、検索曲数を増大させる事が目的である。現実の楽曲セールス数をオリコンチャートより分析した。ある時間間隔で観察した場合、主に「流行」という要素により検索要求頻度の高い楽曲が発生し

ていることがわかる．従って，検索要求頻度の高い楽曲 ID を検索対象としてハードウェア上に実装することを行う．この場合，書換え可能なハードウェア (FPGA) を用いることで検索対象が変化した場合でも書換えによって対応できるようにする．また，この過程において検索部分としては，1 楽曲 ID の検出器を並列に実装するとした．これは検索可能な楽曲数の増大につながる．これによりソフトウェアの部分の検索時間の制限を補償する．

システム全体の構成としては，ソフトウェアとハードウェアそれぞれの優れている部分を効率的に利用するために，楽曲 ID を入力後，一度，ハードウェアで検索を行うとした．その理由としては，検索頻度の高い楽曲検索部分を実装してあるため，とソフトウェアと比較し高速に処理できるためである．より早い段階で楽曲 ID を検索完了する事が検索時間の短縮につながるためである．検索頻度の高い楽曲検索部分において，どの楽曲検出をハードウェア上に実現すべきか，についてはオリコンチャートからの分析によって得られた結果を利用する．これらハードウェアとソフトウェアを組み合わせることで，検索要求の柔軟な変化に対応できるシステムを提案する．

### 3 評価

提案手法の有効性を示すために，得られたソフトウェアのみでの楽曲 ID の検索時間とハードウェアのみにおける検索時間について，それぞれ評価を行う．作成したハードウェアではどの程度並列実装が可能か，を評価する．また，そのときの動作周波数についても評価する．次にこれらを組み合わせた場合の各パラメータを評価する．提案システムではまず始めにハードウェアで検索を行うとした．ここで楽曲 ID が識別でき検索完了となる場合と，次のソフトウェア部分で検索処理を行う場合とに分かれる．システム全体の性能としては，このハードウェアでの識別確率が大いに影響すると考えられる．よって，このハードウェアでの識別確率をオリコンチャートからどう与えるべきか，を評価する．また，更に，このハードウェアでの識別確率が，時間の経過によってどうなるのか，をオリコンチャートから得られたデータから評価する．ハードウェアを書換え可能な FPGA デバイスを用いて構成するので，この書換えにかかる時間を示す．最終的には書換えの頻度とハードウェアの識別確率の関連性を述べた後，どのように与えるかを示す．

### 4 まとめと今後の展望

楽曲ファイルを対象とした本稿では，コンテンツ識別技術のフィンガープリントをテーマに，大規模かつ高速な一致検索をハードウェアとソフトウェアのハイブリッドシステムを用いて実現した．

また現在流通している楽曲を分析した結果コンテンツの流通具合は「流行」という要素に左右される事が分かった．この「流行」という要素を利用し，高速に検索を行うシステムを製作し，挙動を確認できた．これに着目して，再構成可能なデバイスをハードウェア

として用いる事で高速に一致検索を行う手法を確立できた。更に、流行の変化などにハードウェアも柔軟に対応できるようにした。これによってその時期に最頻度の参照データをハードウェアで高速に検出する事が可能である。

通常、ハードウェア上に実装する一致検索システムではハードウェア上の制限から検索データベースサイズが制限されるので、大容量にすることが出来ないが、本研究ではこれをソフトウェアによる柔軟なサポートで解決をした。

100Mbps でリンク状態の回線で 7.2MB の楽曲ファイル転送を行うと楽曲 ID 検出まで 314ms(礒永ら) かかることがわかっている。これをふまえれば、検索余裕時間が 250ms となる。速度の面から言えば、ハイブリッドシステムであるので、システム全体の処理時間に対し処理時間の最もかかる部分が支配的になる。しかしながらハードウェアでの識別率が  $996.59 \times 10^{-3}$  以上で、ソフトウェア検索時間が 100 万曲分を検索するために 73.29sec 以内であるならば設計目標であるシステムの総検索時間を 250ms 以内に収める事が出来る。ハードウェアでの識別率が  $996.59 \times 10^{-3}$  以上必要だとすれば、2007年3月12~2007年3月18日におけるオリコンチャート集計データだと 1~290 位分をハードウェアに実装する必要がある。よって本システムでは検索タイトル数が 290 曲目以上であるならば、実装を行ったハードウェアとソフトウェアによるハイブリッドリアルタイム検索システムが有用といえる。

逆に本実験より得られたソフトウェア検索時間、ハードウェア検索時間を利用し、検索目標時間を 250msec 以内とするならば、オリコンチャートの分析より得られたハードウェアでの識別確率を 0.9 とすれば、ソフトウェア検索時間は 2.5sec 以内でなければならない。このことから、実験によって得られたソフトウェア検索時間について検討すれば、検索可能な楽曲数は実験よりおよそ 35700 曲分とわかる。

さらに、性能を維持するためのハードウェアの書換頻度として、従来手法の LRU を用いる方法と本提案手法であるオリコンチャートを参照する方法を比較した。その結果、LRU 手法では 1.8 分に 1 回と高頻度の書換を要求することに対し、本提案手法では最低でも 50 分に 1 回の書換えで済むことを明らかにした。

ハードウェア上に実装可能かどうかは FPGA の面積的な問題とも関わってきている。また書換時間も並列度を向上させると増加してゆく。これはチップの集積度の向上、マルチチップ運用、ホスト PC の並行稼働など、様々な方法で実現可能といえる。

また本実験で用いたソフトウェアでの検索において、速度やデータベースの格納方法などに現実的ではない部分があり、これは今後の課題となりうる。