フォトアルバムから音楽を生成するプログラムの対話型 GP を用いた最適化

安 藤 大 地† 笠 原 信 一†

写真やフォトアルバムから音楽を作り出す生成的手法やシステムは様々なものが提案されているが，そのほとんどはシステム作者の独断の決定論的手法を実装したものが多い．そのため生成された音楽がユーザのイメージと合わなかったり，どんな写真やアルバムに対しても同じような音楽しか生成できない，といった問題が生じていた．そこで対話型の遺伝的プログラミング (GP) を用いる事で，ユーザのイメージや好みを反映しつつ，フォトアルバムごとに動的に音楽を生成するプログラムを作る手法を提案する．

Interactive Optimization of Music Generation Method from Photo Albums by means of Interactive Genetic Programming

Daichi Ando† and Shinichi Kasahara†

This paper propse a technique of interactive optimization of background music generation method from photo albums. Exists systems that generating music from photos have only static music generating mapping determined by developer. Thus there are estrangements between actural music output of these systems and user's feelings for inputed photos. We propse interactive optimization system for music generating mapping to prevent such estrangements.

1. 導　　入

映像と音楽のリアルタイムインタラクションについては，映像のリアルタイム生成が可能になった 90 年代後から盛んに試みられてきた．映像だけではなく写真展に合わせた生演奏なども現代芸術の分野では多く行われている．また近年の iPhone などのカメラつきスマートフォンの登場で，フォトアルバムと他のメディアとの連携についても注目されるようになってきた．

写真やフォトアルバムから音楽を作り出す生成的手法やシステムは様々なものが提案されているが，そのほとんどはシステム作者の独断の決定論的手法を実装したものである．写真展に付属した生演奏などアーティストがアート作品としてその場で行うケースは除けば，多くの場合でシステムに準備されている固定された写真を音楽に変換するマッピングを用いている．

そのため生成された音楽がユーザのイメージと合わなかったり，どんな写真やアルバムに対しても同じような音楽しか生成できない，といった問題が生じていた．写真に対するイメージは人それぞれであるが，写真の情報から固定されたマッピングで音楽を生成した場合，ユーザのそれらの写真に対するイメージを全く無視していることとなる☆．例えば「奇麗な 3DCG」を集めたフォトアルバムに固定された変換マッピングのみを適用し音楽を生成すると，ユーザの写真に対する色合いや構成などから受ける，もしくは演出したいイメージを無視して，特定の感情を抱かせてしまう音楽ばかり出力されてしまう可能性が高い．

そこで対話型の遺伝的プログラミングを用いる事で，ユーザのイメージや好みを反映しつつ動的に音楽を生成するプログラムを作り，これを変換マッピングに挿入する手法を提案する．提案手法を用いることで自分のフォトアルバムの少数の写真を元にあらかじめ自分の好み，もしくはイメージに沿った BGM を生成する変換マッピングを作っておくことで，フォトアルバム内の他の写真から色合いや写真ごとの類似度の変化などの情報からユーザ好みの音楽を作り出すことが可能になる．

2. 対話型進化論的計算

対話型進化論的計算 (Interactive Evolutionary

† 首都大学東京システムデザイン学部
Faculty of System Design, Tokyo Metropolitan University

☆ 普通のスナップ写真のアルバムに対しては，その写真が撮られた時の思い出に対する感情が多くを占め，写真の色合いなどから音楽を生成する変換マッピングは役に立たない事は容易に想像できるため，このケースは除く

Computation, IEC) とは，遺伝的アルゴリズム (Genetic Algorithm, GA) や遺伝的プログラミング (Genetic Programming, GP) などの進化論的計算の評価関数を人間に置き換えたものである．具体的には，生成された個体の評価を行う際に，人間が個体を直接見て (聴いて)，その個体の評価値を決定する．提案手法では，GP で生成したプログラムが生成する音楽をユーザが聴き善し悪しを決める事が評価となる．

IEC は評価関数を設計する事が難しい問題，例えばアート作品やデザインの分野，使用する個人によって最適解が異なる個人使用機器のプログラムの最適化，さらに多目的最適化 (Multi-Objective Optimization, MO) を必要とする基礎設計などで有効である．

しかしながら，評価関数となる人間の肉体的，精神的負担が非常に大きい．ユーザインタフェースや進化プロセスの工夫により，ユーザ負担をどれだけ減らせるかが重要となる．

3. 提 案 手 法

3.1 概　　要

ユーザは，自分のフォトアルバムを表示させながら，そこから音楽を生成するプログラムを，実際に生成される音楽を聴きながらプログラムの交配と音楽の生成を繰り返して対話的に最適化していく．

提案手法ではフォトアルバムは合計 9 枚 (1 ページに 3x3 が表示されているフォトアルバムのイメージ) を入力に取り，4 小節分の音楽フレーズをリアルタイムで生成し，これがループされる．3x3 のページを 2 枚連続させ 8 小節単位のループにすることも可能である．

また，最適化後のプログラムに対して，入力となっているフォトアルバムの写真をリアルタイムに変更していく (アルバムのページをめくるなど) の操作を行うと，リアルタイムで生成される音楽も変わっていく，という状態となる．この写真の入れ替えは最適化中にも自由に行えるため，色々な写真を入れ替えながら実際に生成される音楽のモニタリングを行う事も出来る．

ユーザインタフェースは，時系列メディアの IEC に特有の問題を解決するため，個体をアイコンで表現し，個体アイコンの移動を伴う Simulated Breeding の二値評価を導入する．

3.2 基礎マッピングと対話 GP の二重構造

画像情報を入力に取り GP で音楽全ての要素を生成しようとすると，一般ユーザを対象とした対話型ではユーザ負担が大きく最適化が困難である．そこで提案手法では，音楽の要素を生成する「基礎マッピング部」

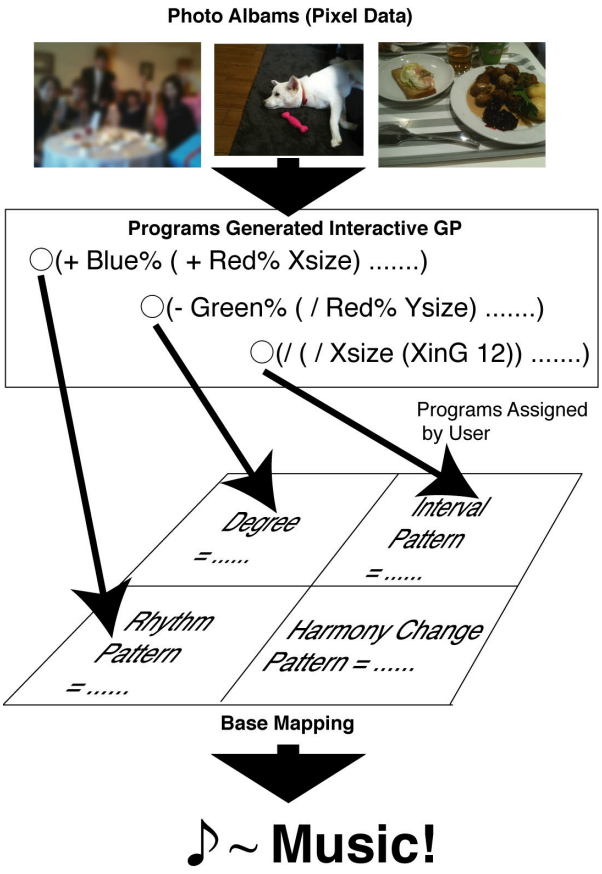

図 1　基礎マッピング部分と対話型 GP 部分の二重構造

を準備し，画像の様々な数値を出力するプログラムを対話型 GP で生成するという，二重構造をとる．図 1 に提案手法の二重構造を示す．

この手法により，どんな画像からでも最低限一般ユーザの聴取に足る音楽を生成することが可能であり，さらにユーザの好みに応じてそれを低負担で変更可能，というシナリオをユーザに提案することができる．

GP の関数ノードは通常の整数のみの数値演算遺伝子表現を用い，個体そのものは整数を出力するが，マッピングにアサインする際に出力を 0〜1 へ正規化する．基礎マッピングでは音楽フレーズの生成ルールをそれぞれ 4 つ用意しておき，ユーザの好みでアサインされた GP の個体の出力を使ってそれぞれのルールを用いて音楽フレーズを生成する．

3.3 基礎マッピング部

提案手法は，フォトアルバムの BGM「一般的に耳障りがよく」「オシャレっぽい」4 拍子イージーリスニング調のループ音楽を生成することを目的とする．

アルバム一つ (9 枚もしくは 18 枚表示) からハーモニートラック，リズムトラック，メロディトラックのいずれかを生成する．これは複数のアルバムが同時に

表示され，そのセッションで一つの音楽が生成されることを前提にしているからである．

それぞれのトラックは，GP で生成したプログラムをそれぞれアサインする生成ルールを 4 つ備えており，ユーザが後述するインタフェースでプログラムをアサインしていく．生成ルールを 4 つに限定したのは，後述のアサイン作業でユーザが最適化の負担を感じることなく音楽的発見を楽しむためには，探索しなければ行けない生成ルール数が多すぎたり少なすぎるのでは不都合であるためである．

例として示すハーモニートラックは，以下の 4 つの生成ルールで構成される．

(1) ハーモニーの基音となるディアトニック度数の 4 つの組み合わせパターンの選択

(2) (1) の基音から和音を生成する際の各構成音インターバルのパターンの選択

(3) (1) と (2) から作られる和音をどのように切り替えていくかのハーモニー変更パターンの選択

(4) (1) と (2) から作られる和音のアルペジオフレーズのリズムパターンの選択

これらの生成ルールの選択パターンはイージーリスニングとして一般的に耳通りがいいものをそれぞれ複数用意する．例えば (1) のパターンを作るときは，一番シンプルな方法では 7 度の音から 4 つを重複可能で選ぶので単純に 2400 通りがありうる．これを 0〜1 の間に等間隔で振り分け生成ルールを用意した．(2) でインターバルにより和音を構成するルールには，テンションノートを盛り込んで「ちょっとオシャレ」な和音を作り出している．

リズムトラックの生成ルールは，提案手法が 4 拍子固定であることを利用し，楽器 4 種のリズムパターンを 4 つの生成ルールに割り振る．

また，メロディトラックは，音程のパターンを生成するルール 3 つとリズムパターンを生成するルール 1 つから成り立っている．

3.4 GP によるプログラム生成

GP 部分は通常の算術演算のみシンプルな GP を用いる．

実際には画像のピクセルデータの分析結果を終端ノードとして取り，通常の整数の算術演算を関数ノードとして用い，整数値を出力する．これを 0〜1 に正規化したものがマッピングされる．表 1 に主要な終端ノードを示す．添字には，現在表示されている写真の中で何番目の写真か，という情報が入る．終端ノードが取る数値は 0〜255 の範囲内である．

表 1 GP の終端ノードの例．添字は写真のインデックスが入る

R%a	画像全体の R 要素の割合 (G と B も同様)
Xsizea	画像の Xsize
Ysizea	画像の Ysize
SimilCol	9 枚の画像の色合いの類似度
ClustRa	R 要素の塊の大きさ (G と B も同様)
ClustWa	RGB ともに高い塊の大きさ (K も同様)

図 2 ユーザインタフェース全景．左部にフォトアルバム部分，右部に IEC 部分が表示されている．

3.5 ユーザインタフェースと IEC の進化プロセス

図 2 に，ユーザインタフェース全体を示す．左部にフォトアルバム部分，右部に IEC 部分が表示されている．フォトアルバムは前述の通り 3x3 の 9 枚 (4 小節相当) と縮小し 3x6 の 18 枚 (8 小節相当) を表示可能であり，左下の “9pics ←→ 18pics” というボタンで切り替えられる．その右側の “NEXT PAGE” ボタンで，フォトアルバム次のページ (次の 9 枚もしくは 18 枚) へ進める事が可能である．

IEC のユーザインタフェース部として，Dahlstedt らの Mutasynth[1] を元に，著者らの時系列メディア用 IEC のインタフェースの改善[2] を盛り込んだものを提案する．図 3 に図 2 の IEC 部の拡大を示す．

提案ユーザインタフェースでは，各個体が染色体 (ここでは S 式) から作り出されたヒモ状のアイコンで表示されている．上から “Assign Area”，“Parent Area”，“Variation Area”，“Storage Area”，“Initialize Area” の 5 つのエリアが存在し，個体をドラッグアンドドロップで移動 (コピー) することが可能である．Assign Area の 4 つの個体スペースは前述の基礎マッピングの四つの基礎マッピングに対応しており，任意のプログラムを容易にアサイン可能なインタフェースとなっている．

最適化の手順は以下のようになる．

(1) Initialize Area にある Inialize ボタンを押し，

Initialize Area に 6 つのランダム生成の初期個体を登場させる．個体は Assign Area にドラッグアンドドロップで移動させることで聴取することができる．

(2) 任意の二つの初期個体を Parent Area に移動させ，Crossover ボタンを押すと，Variation Area に 8 つの子世代が誕生する．

(3) Variation Area の個体を Assign Area に持っていき評価しながら，気に入った個体を再び Parent Area に入れるなどして最適化を進める．また行き詰まったら Initailize Area で再びランダム生成初期個体を作り，進化プロセスに挿入する．

(4) 最適化途中の一時的な個体の置き場として Storage Area を用いる．

この手順は，通常の進化計算の最適化プロセスとは異なり二値評価を繰り返す Simulated Breeding に近いものとなる．時系列メディアの最適化においては評価をするための聴取そのもののユーザ負担が大きく，また「面白い個体を発見する」行為として捉えることが有効である．そのためにこのような個体アイコンの移動を伴うシンプルな評価付けが有効である事が著者らの研究[2] で示されている．

3.6 システム構成

プロトタイプとして作成したシステムは三つのユニットで構成されている．一つは写真を読み込む解析を行う機能を持つユーザインタフェース部分，GP のエンジン部分，そして基礎マッピングを内蔵し実際に音楽を発音する部分である．ユーザインタフェース部分は現在は Processing JAVA で構成されており，GP エンジンは将来の移植を考え C++，音楽を発音する部分は SuperCollider とソフトウェア MIDI 音源で作られている．それぞれのユニットは OpenSound Control で接続されている．

また，音楽発音部分一つに対して，ユーザインタフェース部分と GP エンジン部分を複数接続する事が可能になっており，いくつかのアルバムを持ち寄り同時に表示，探索しながら複数人数でセッションを行うことも可能である．

4. ま と め

本稿では，フォトアルバムの BGM を生成するプログラムの対話型 GP を用いた最適化の提案を行った．従来型のシステムではシステムにより提示された画像から音への変換マッピングしか用いることができなく，ユーザのフォトアルバムに対するイメージと生成された音楽へのイメージが乖離してしまう．そこで生成ルールを多数もうけ画像情報から生成ルールへのマッピングを GP により最適化する事で，ユーザのイメージに沿った BGM を生成する事が可能になった．また対話型のインタフェースでは，時系列メディアの最適化のためのインタフェースの改善を行うことにより，ユーザ負担の軽減を行った．

将来の展開として，インタフェース部分を iPad などの写真ストレージ機能を持っている携帯デバイスに移植することを考えている．

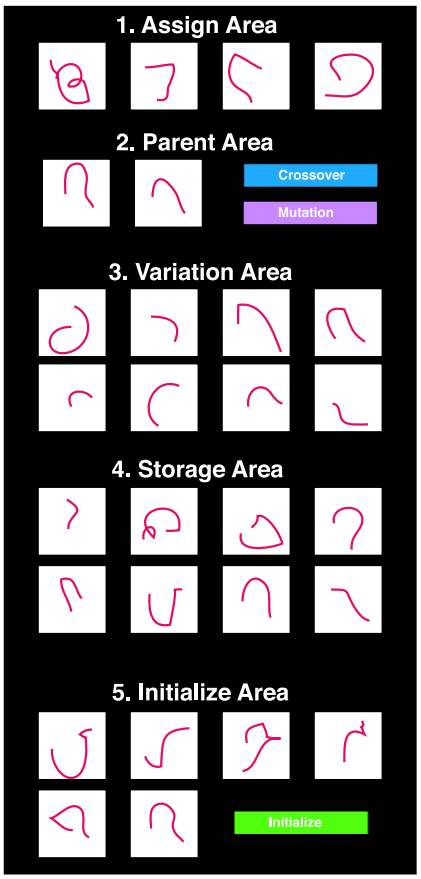

図 3 IEC 部分のユーザインタフェース．各個体はヒモ上のアイコンで表示される．アイコンの移動で音楽生成へのアサインなどを行う．

参 考 文 献

1) Dahlstedt, P.: A MutaSynth in parameter space: interactive composition through evolution, *Organized Sound*, Vol. 6, pp. 121–124 (2004).

2) 安藤大地，稲田雅彦，丹治　信，伊庭斉志：能動的音楽聴取インタフェースの作曲支援 IEC への取り込み，第 73 回情報処理学会音楽情報科学研究会，pp.1–6 (2007).