E-034

歌詞特徴を考慮したWeb画像と楽曲同期再生システムの提案

A Proposal for Synchronized Web Image and Music Playback System using Lyrics

舟澤 慎太郎† 石先 広海‡ 帆足 啓一郎‡ 滝嶋 康弘‡ 甲藤 二郎†
Shintaro Funasawa Hiromi Ishizaki Keiichiro Hoashi Yasuhiro Takishima Jiro Katto

1. はじめに

映画やテレビ番組，ミュージックビデオでは，映像と音楽を効果的に組み合わせることにより，それらを単独で視聴する以上のインパクトを生み出している．このような相乗効果を，普段音楽を視聴する際にも適用することで，より印象深い音楽体験が実現できると考えられる．

そこで本稿では，音楽を主体にそれに合う画像を Web 上から自動検索し，スライドショーとして音楽と同期再生するシステムを提案する．音楽に合う画像の検索は歌詞特徴を基に行う．しかし，単に歌詞に出現する単語をクエリとして検索を行うだけでは適切な画像を取得することは難しい．本稿ではこの点に着目し，音楽に合う画像を検索するためのクエリ選定法について検討する．そして，被験者評価実験により本手法の有効性を検証する．

2．関連研究

本研究のように，楽曲の歌詞に基づき Web 上から画像を検索し楽曲と共に再生するシステムが提案されている[1][2]．これらの研究では，歌詞を基に画像を検索した後，その結果から如何にして最適な画像を選定するか，という処理に重きを置いており，画像検索に与えるクエリの選定という点では，単純な処理しか行っていない．しかし，歌詞に出現する単語にも画像として表現する上で重要なものとそうでないものが存在するため，クエリを適切に選定しない場合，意図しない画像が検索されることがある．また上記文献では，画像検索の結果から最適な画像を選定するために，何らかの画像処理を行っている．

以上の点を考慮すると，あらかじめ画像検索に与えるクエリを適切に選定することで，不適切な画像の取得を避けることができ，更に，後に行う画像処理に要する処理時間の軽減にもつながると考えられる．

3. Web画像と音楽の同期再生システム

3.1 システム概要

図 1 にシステムの概要を示す．前提として，再生対象となる楽曲の音源，歌詞情報及び，同期情報がシステムの DB 内に格納されているとする．同期情報とは，歌詞の各行の楽曲中における開始と終了の時間を示す情報である．

システムは入力としてユーザから楽曲の指定を受けると，“クエリ選定部”において，指定された楽曲の歌詞を用いて各行ごとのクエリ候補単語を抽出する．次に“画像検索部”において，クエリ候補単語を基に画像 DB から画像を検索する．本研究では，画像 DB として写真共有サイト Flickr[3]を用い，画像に付与されているタグを用いて検索を行う．そして“画像選定部”において，画像検索結果の中から最適な画像を選定する．最後に“同期再生部”において，検索した画像を楽曲と共に再生し，ユーザに出力する．

† 早稲田大学 理工学術院
‡ 株式会社 KDDI 研究所

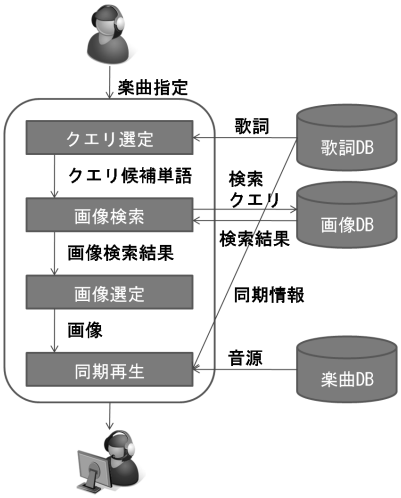

図 1　システム概要

3.2 提案手法

（1）クエリ選定部

歌詞の各行ごとに画像検索に与えるクエリ候補単語を抽出する．ある行におけるクエリ候補単語は，1: 行に出現する名詞，2: 行を含む段落に出現する名詞，3: 楽曲全体の推定印象語，にて構成される．3 については，文献[4]にて行った歌詞に基づく楽曲分類の季節， 時間帯， 天候カテゴリにおける分類されたラベル(夏，朝，等)を用いる．

検索対象の行に出現する名詞集合を $\mathbf{N}_1$，行を含む段落に出現する名詞及び，楽曲全体の推定印象語から成る集合を $\mathbf{N}_2$ とする．更に $\mathbf{N}_1$ ，$\mathbf{N}_2$ において，DF(document frequency)又は UF(user frequency)が閾値以下の要素を排除する．ここで DF(**W**)とは，単語集合 **W** の要素全てを用いて Flickr で AND 検索をした際の，検索された画像数を示し，UF(**W**)とは，検索結果におけるユニークな画像投稿者数を示す．閾値は DF を 40，UF を 10 に経験的に設定した．こうして，歌詞の各行ごとのクエリ候補単語を抽出する．

（2）画像検索部

クエリ候補単語から最適な組合せを選び，画像を検索する．本手法は“行の出現単語を最優先する”，“多くの単語を用いてクエリを構成し，より詳細に絞り込みを行う”，“タグとして付与されやすい単語(UF の高い単語)を優先する”，という考えに基づいている．DF ではなく UF を重視するのは，Flickr では少数のユーザが大量の写真に同一タグを付与し，DF が不当に高くなる場合があるためである．以下に処理の流れを示す．

1. $\mathbf{N}_1$ のべき集合 $\mathbf{P}(\mathbf{N}_1)=\{\mathbf{W}_{1,1}, \mathbf{W}_{1,2}, \ldots, \mathbf{W}_{1,x}\}$において，$DF(\mathbf{W}_{1,i}) > 1$ を満たし，かつ，$|\mathbf{W}_{1,i}|$ が最大となるような $\mathbf{W}_{max}$ を選出する．ここで $|\mathbf{W}|$ とは，単語集合 **W** を構成する要素数(名詞数)を示す．$\mathbf{W}_{max}$ が複数ある場合，$UF(\mathbf{W}_{max})$が最大のものを選出する．こうして選出した $\mathbf{W}_{max}$ を行クエリ集合 $\mathbf{Q}_{line}$ とする．

図 2　デモアプリケーション画面

2. $\mathbf{N}_2$ のべき集合 $\mathbf{P}(\mathbf{N}_2)=\{\mathbf{W}_{2,1}, \mathbf{W}_{2,2}, \ldots, \mathbf{W}_{2,y}\}$ の各要素に $\mathbf{Q}_{line}$ を加えた集合 $\mathbf{P'}(\mathbf{N}_2)=\{\mathbf{W}_{2,1}\cup\mathbf{Q}_{line}, \mathbf{W}_{2,2}\cup\mathbf{Q}_{line}, \ldots, \mathbf{W}_{2,y}\cup\mathbf{Q}_{line}\}=\{\mathbf{W'}_{2,1}, \mathbf{W'}_{2,2}, \ldots, \mathbf{W'}_{2,y}\}$ において，**1** と同様の処理により，$\mathbf{W'}_{max}$ を選出する．

3. $\mathbf{W'}_{max}$ の要素全てを用いて Flickr にて AND 検索を行う．

以上の処理を，歌詞の各行に対して行う．

（3）画像選定部

検索出力結果の中から最適な画像を決定する．現在は Flickr における“most interesting”指標でランクが最も高いものを選定している．但し，同楽曲内で同じ画像は選定しない．こうして，歌詞の各行に対応する画像を決定する．

（4）同期再生部

楽曲と歌詞の同期情報を用いて，検索した画像と楽曲を同期再生し，ユーザに提示する．出力のインターフェースは図 2 にようになっており，画像だけでなく，対応する歌詞や楽曲のメタ情報も共に表示する．

3.3　比較手法: TF*IDF によるクエリ選定

もう一つの手法として，TF*IDF によるクエリ選定法を検討する．TF*IDF は楽曲を特徴付ける歌詞中の単語の重要度を表現する指標である．なお，TF*IDF は市販の CD から収集した J-POP 楽曲 3062 曲を基に算出した値を用いた．以下に処理の流れを示す．

まず“クエリ選定部”において，3.2 と同様 $\mathbf{N}_1$ を抽出する．そして“画像検索部”では，$\mathbf{N}_1$ の全要素を用いて AND 検索を行い，$DF(\mathbf{N}_1) = 0$ ならば，$\mathbf{N}_1$ の中で最も TF*IDF の低い要素を排除し，再び検索を行う．この処理を画像が取得できるまで繰り返し行い，$|\mathbf{N}_1| = 0$ となっても画像が取得できなければ，前の行のクエリを用いて検索を行う．“画像選定部”と“同期再生部”では 3.2 と同様の処理を行う．

4. 評価実験

4.1　実験内容

クエリ選択手法の有効性を検証するため，被験者による評価実験を行った．被験者は，3.2 で示した提案手法と 3.3 で示した比較手法の 2 つの手法にて生成したスライドショーを楽曲と共に視聴し，両手法について，全体評価と個別評価を行う．全体評価として，楽曲全体で同期再生した画像群の適切性を，5:とても楽曲に合っていた～1:全く楽曲に合っていなかった，の 5 段階で評価する．また，個別評価として，歌詞の行に対する画像の適切性を評価する．具体的には，視聴者が適していると判断した画像を選択してもらうことで評価する．

なお，被験者は大学生 20 名で，楽曲データは J-POP 楽曲 10 曲を用い，1 曲につき 10 名分の評価を収集した．

表 1　各手法の画像適合率と全体評価平均値

楽曲 No	画像適合率		全体評価平均値	
	提案手法	比較手法	提案手法	比較手法
38	**55.9**	42.6	**3.9**	**3.9**
88	45.5	**47.6**	3.9	**4.4**
153	**57.9**	37.9	**4.2**	3.3
163	**38.9**	28.6	**3.5**	3.2
452	**50.0**	20.9	**3.6**	2.4
平均	**48.3**	33.9	**3.7**	3.3

4.2　実験の流れ

1. 1 つ目の手法にて生成した楽曲スライドショーを視聴し，この手法に対する全体評価を行う．

2. 同じ楽曲について，もう一方の手法で生成した楽曲スライドショーを視聴し，同様に全体評価を行う．

3. 評価ページにおいて，視聴した楽曲について両手法の個別評価を行う．

以上の処理を提示された 5 曲全てについて評価してもらう．なお，両手法においての順序効果は考慮してある．

4.3　実験結果

表 1 に評価結果の一部を示す．表中の画像適合率は，個別評価において楽曲に適していると判断された画像の割合を示し，全体評価平均値は全体評価における被験者 10 名の平均した評価値である．全体の傾向として，提案手法の方が，両評価値共に良い結果となっている．

比較手法の方が優れていた楽曲 No.88 に関しては，DF, UF による閾値処理により本来用いるべきクエリが排除されていた，クエリの拡張自体は適切だがそれにより Flickr 側の適切でないタグが付与されている画像が取得されていた等の特徴が見られた．ここでいう適切でないタグとは，画像投稿者が自分の画像を検索されやすくするために付与するタグや，個人の主観により付与したタグ等，画像の客観的な内容とは関係のないタグのことである．

この問題の対策として，選定したクエリで検索し取得した画像群に対し，画像自体の特徴を考慮することにより，画像が本当に楽曲の内容に適しているかどうかを判断する処理が必要であると考えられる．

5. まとめ

本稿では，歌詞特徴に基づく Web 画像と楽曲の同期再生システムの実現のための，画像検索に与えるクエリ選定法の検討を行った．また，被験者評価実験において，提案手法の有効性を検証した．今後は，検索した画像群から最適な画像の選定，楽曲の音響的特徴の考慮，画像の切り替えタイミング等の検討を行う予定である．

参考文献

[1]S.Xu *et al.*: “Automatic Generation of Music Slide Show using Personal Photos”,IEEE ISM 2008, pp.214-219

[2]R.Cai *et al.*: “Automated Music Video Generation Using Web Image Resource”, IEEE ICASSP 2007, Vol.2, pp.737-740

[3]Flickr: http://www.flickr.com/

[4]舟澤ほか: “歌詞の印象に基づく楽曲検索のための楽曲自動分類に関する検討”, 第 71 回情処全大, 5R-2, 2009