

異種メディア相互検索システムに関する検討

- システムの評価 -

林坂 享祐[†] 梶川 嘉延[†] 野村 康雄[†]

[†] 関西大学工学部 〒564-8680 大阪府吹田市山手町3-3-35

E-mail: †{rin,kaji,nomura}@joho.densi.kansai-u.ac.jp

あらまし 本稿では、感性を用いた異種メディア相互検索システムに関する検討を行う。我々はこれまでに音楽、画像における感性検索システムの構築を行っている。前回の報告において従来の感性を用いた異種メディア相互検索システムの問題点を踏まえて、それら両者の検索システムの結び付けを行った。その結果、異なる印象を持った2つのメディアが一致することを示した。しかし、システムの評価が不十分であったので、今回は新たに行ったシステムの評価の結果について報告する。

キーワード 感性検索システム, 異種データ結び付け

A Study on Inter-Retrieval Systems among Different Media

- System Evaluation -

Kyosuke RINSAKA[†], Yoshinobu KAJIKAWA[†], and Yasuo NOMURA[†]

[†] Faculty of Engineering, Kansai University Yamate-cho 3-3-35, Suita-shi, Osaka, 564-8680 Japan

E-mail: †{rin,kaji,nomura}@joho.densi.kansai-u.ac.jp

Abstract In this paper, we study a KANSEI inter-retrieval system among different media. We have already proposed KANSEI retrieval systems for music and pictures. In the last report, we have combined the KANSEI retrieval systems for music and pictures in order to solve some problems of conventional KANSEI inter-retrieval systems. Some experimental results demonstrate that two media with different impressions can be related. However, since the system evaluation has been insufficient, this paper shows new evaluation results of the proposed KANSEI inter-retrieval system.

Key words KANSEI Retrieval Systems, Different Data Connecting

1. ま え が き

近年、計算機技術の発達により、マルチメディア情報をコンピュータで扱う事が技術的にもコスト的にも特別な事ではなくなってきた。ここで言うマルチメディア情報とは、映像と音響が中心である。マルチメディア機器の普及により、演出効果に関する専門家でないエンドユーザが、マルチメディア作品を自分で創る機会が増加してきている。マルチメディア作品は画像や音楽といったマルチメディア情報を効果的に用いて相手に訴えかける事を目的とするので、それら演出効果の品質は作品の良し悪しを決める重要なものである。ところが、エンドユーザにとって、効果的にマルチメディア情報を用いた高品質の作品を制作する事は非常に困難である。従来、これらの作品は、監督(ディレクター)、デザイナー、音楽家等の視覚・聴覚演出効果に関する専門家が、多くの資金と時間を費やし、共同作業を

する事によって創られてきた。

そこで、画像に見合う音楽を自動的に選定する事ができるならば、作品の制作に費やしていた多くの資金と時間を軽減する事ができる。またエンドユーザも、効果的にマルチメディア情報を用いた高品質の作品を制作する事ができる。この事は、音楽に見合う画像を自動的に選定する場合についても同様の事が言える。

このことに関連して画像・音といったマルチメディアデータを内容に基づいて検索したいという要求があり、さまざまな研究が行われている。それらの研究は、画像なら画像といった単一のメディアを扱ったものである[1], [2], [3]。ここで、上記で述べたマルチメディア作品を作成したい場合、そのエンドユーザが各分野に精通していないとすれば、効果的な演出を選ぶことは困難である。このような場合には、異なるメディア間にまたがって有効なデータを取り出すことが要求される。

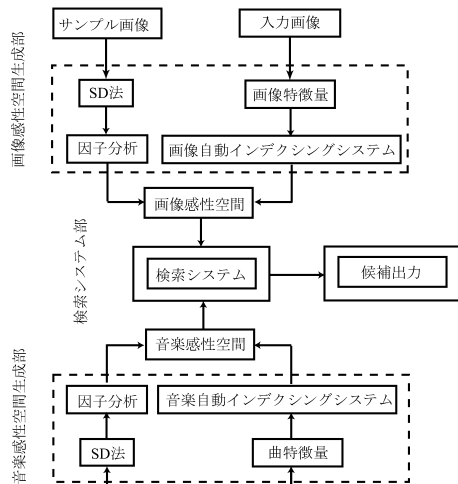


図1 システム概要

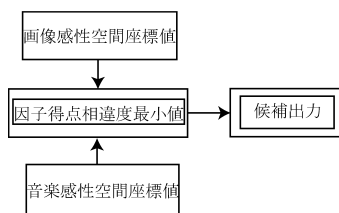


図2 検索システム部概要

そこで本研究では、異種メディア間における相互検索を可能にするシステムの構築を行っている。先行研究として文献 [4], [5], [6] が報告されている。文献 [4], [5] では、画像の自然性と音楽の自然性というように同様に解釈した主因子を共通するものとして結び付けを行っている。しかし、例えば、夜景の町並み（暗い画像）と楽しい音楽というように画像と音楽の印象が異なっても見合うものもあるにも関わらず、印象の一致度により検索しているという点で問題があると考えられる。また、文献 [6] では、動画と音楽を同じ感性語で評価を行い、それらの一致度を求める事により検索を行っている。しかし、動画、音楽を評価する際、それらの感性語は言葉としては同じでも各々の印象において意味合いが必ずしも一致するとは限らないという点で問題があると考えられる。そこで本研究では、画像と音楽各々に適した感性語を用意し評価を行っている。前回の報告 [7] において、画像と音楽を結び付けるために行った印象の関係に関する調査結果、ならびにその結果を利用して構築した相互検索システムの評価を行った。前回では、ランダムに選んだ画像において、本システムが選出した結果のみを用いて評価を行っていた。しかし、人間は心理的に画像と音楽を合わせようとすることから、画像と音楽が合わないものは無いのではないかとこの指摘を受け、システムの評価が不十分であった。本稿では、この指摘を踏まえて新たに行ったシステム評価の結果を報告する。

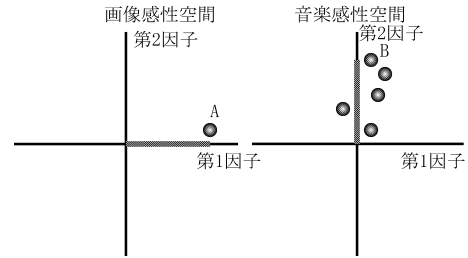


図3 候補選択例

2. システム概要

ここでは、本研究で構築を行っている検索システムの概要について示す。異種メディア間における相互検索を可能にするシステムに関する先行研究として文献 [4], [5], [6] が報告されている。文献 [4], [5] では、例えば、画像の自然性と音楽の自然性というように同様に解釈した主因子で結び付けを行っている。また、評価実験に基づいてメディアデータベースを構築し、検索を行っている。しかし、新たに画像などをデータベースに登録するたびに評価実験を行うのは実用的ではない。

そこで本研究では、自動的にメディアの印象を求め、より実用的なシステムの構築を目指している。

検索システムの構成図を図1に示す。図1を見てもわかるようにこのシステムは大きく分けて「画像感性空間生成部」、「音楽感性空間生成部」、「検索システム部」の三つの部分から構成されている。

画像感性空間生成部は、画像の印象（明るい、暖かいなど）の度合いによって、画像印象の心理的相関を表す画像感性空間における画像の位置を決定する部分である。SD法を用いて画像印象を求め、因子分析を用いて感性空間上の位置を決定する。ここで、画像を登録する度にSD法による評価実験を行うのは手間がかかる。そこで、評価実験を行っていない画像の画像感性空間座標値を自動的に求める「画像自動インデクシングシステム」を用意する。画像自動インデクシングシステムは画像から抽出することのできる特徴量（色情報 etc.）を用いて画像の感性空間座標値を求めるシステムである。また、音楽感性空間の生成方法は、画像感性空間の場合と同様である。我々はこれまでに音楽感性空間生成部に関して文献 [8] で提案している。

検索システムは、画像感性空間と音楽感性空間を結び付け、画像に見合う曲、曲に見合う画像を実際に検索する部分である。検索システム部の概要を図2に示す。印象が結び付いているそれぞれの因子得点の差（距離）をとり、差が最も小さいものを異種メディアに見合う候補とする。例を図3に示す。画像感性空間においてある画像Aが第1因子に対して関連が強く、図に示す因子得点を持っているとする。そして、画像の第1因子と音楽の第2因子が関係があるとすれば、画像Aの因子得点と音楽の第2因子における因子得点の差が最少のものを画像Aに見合う音楽と考える。この場合は音楽Bが画像Aに見合う音楽となる。

画像には写真（風景）、音楽にはピアノ曲（ジャズ、クラシック

表 1 画像に用いた感性語

にぎやか・活気のある	静かな・寂しい
暖かい	冷たい・涼しい
陽気・明るい	陰気・暗い
さわやか	鬱陶しい・重苦しい
雄大	小ぢんまり
重々しい・威厳のある	貧弱・みすぼらしい
モダン・現代的	レトロ・伝統的
沈静的・穏やか	躍動感・迫力のある
派手・豪華	地味・質素
潤いのある	渴いた
自然な	不自然な

表 2 音楽に用いた感性語

明るい	暗い
重い	軽い
かたい	やわらかい
安定	不安定
澄んだ	にごった
滑らか	歯切れのよい
激しい	穏やか
厚い	薄い

表 3 画像における主因子

主因子	感性語 (解釈)
第 1 因子	陽気・明るい - 陰気・暗い (明暗性)
第 2 因子	自然な - 不自然な (自然性)
第 3 因子	暖かい - 冷たい (冷暖性)
第 4 因子	雄大 - 小ぢんまり (景観性)
第 5 因子	重々しい - 貧弱 (威厳性)
第 6 因子	沈静的 - 躍動的 (静動性)

表 4 音楽における主因子

主因子	感性語 (解釈)
第 1 因子	明るい - 暗い (明度)
第 2 因子	激しい - 穏やか (力量性)
第 3 因子	安定 - 不安定 (安定性)
第 4 因子	重い - 軽い (質量性)
第 5 因子	滑らか - 歯切れのよい (躍動性)

ク, etc.) を用いて感性空間を構築している。画像に用いた感性語は文献 [9] を参考にし、表 1 に画像に用いた感性語を示す。音楽についての感性語は表 2 の通りであり、文献 [8] と同様である。画像感性空間を構築する際に、被験者に表 1 に示す感性語に関して思い浮かべる風景のアンケート調査を行い、その結果に基づきデータベースを構築したので本システムにおいて各感性語に見合わない画像はないと考えられる。また、因子分析によって抽出された主因子を表 3、表 4 に示す。表 3 は画像における主因子、表 4 は音楽における主因子を示す。前回の報告 [7] において、画像と音楽を結び付けるために行った印象の関係の結果を表 5 に示す。今回は表 5 の結果を用いて構築したシステムの評価を行う。

表 5 画像を対象とした適合実験結果

画像	音楽
陽気・明るい	安定, 明るい
暗い	暗い, 重い
不自然な	重い
自然な	穏やか
暖かい	明るい, 安定
冷たい	重い, 暗い, 滑らか
雄大	滑らか
小ぢんまり	暗い, 不安定
重々しい	重い, 暗い, 滑らか
貧弱	滑らか
躍動感	重い
沈静	穏やか



図 4 評価実験画面

3. システム評価

3.1 評価実験

表 5 の結果を用いて構築したシステムの評価を行う。前回の報告 [7] で指摘されたように人間は心理的に画像と音楽を合わせようとすることからこの分野の評価は容易ではないが、次のような評価法を考えた。評価を 5 段階評価で行い、5: 合っている, 4: どちらかといえば合っている, 3: どちらでもない, 2: どちらかといえば違う, 1: 違うとする。つまり、評価 4 は心理的に画像と音楽を合わせたと考える。また、評価システムは、提案法を用いて構築したシステム (システム 1)、ランダムに音楽を選択するシステム (システム 2)、画像と音楽が全く適合していないシステム (システム 3) の 3 つのシステムを用意する。画像と音楽が全く適合していないシステムにもかかわらず適合したという評価がある程度得られたなら、人間は心理的に画像と音楽を合わせようとする傾向が強いことがわかる。システム 3 はこれまでの画像および音楽の評価実験を参考にし、実際に自分で視聴して設定した。なお、実験は因子分析によって求めた 6 因子を代表する画像 (12 枚) に絞って行う。

実験は、図 4 に示すダイアログにおいて画像を選択する。そして、図 5 に示すダイアログにおいて評価を行う。被験者は 20 名で、順序効果を考慮するためシステムを数種類用意した。また、システムの内容については被験者に一切の情報を与えなかった。

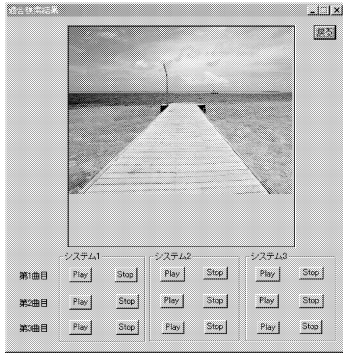


図5 検索結果画面

表6 適合評価

適合度	system1		system2		system3	
	評価数	割合 [%]	評価数	割合 [%]	評価数	割合 [%]
5	75	31.3	51	21.3	8	3.3
4	96	40.0	88	36.7	47	19.6
3	53	22.1	52	21.7	50	20.8
2	15	6.3	39	16.3	79	32.9
1	1	0.4	10	4.2	56	23.3

3.2 実験結果

3.1の条件で行った実験結果を表6, 表7に示す. 表6は240試行における適合評価結果, 表7は各被験者における12試行の評価の平均である. 表6のシステム1, 2の結果より評価4の割合が多く, 心理的に人間は画像と音楽を合わせようとする傾向も見られる. このことについて結果の詳細を見てみると明るい画像, 暗い画像などわかりやすいものは良好な適合評価が得られているが, 特に, 建造物などの画像に対しては合わせようとする傾向が強いことがわかった. また, 被験者は評価5や1のように決まった評価ではなく, 心理的にあいまいな評価を選びやすいことも挙げられる. しかし, この心理的要因を考慮した上で表6, 表7をみてもシステム1は, 他の2つのシステムに比べ, 良好な結果が得られていることがわかる. また, システム3の結果より本システムのデータベースにはどのような画像でも適合するような音楽だけで成り立っているものでないこともわかる.

次に, システムの有意差を χ^2 検定により求める. システム1と2, システム2と3およびシステム1と3の評価5に関して有意差を求めた結果, いずれにおいても0.05を下回り有意差が存在した. よって, 提案システムはランダムシステムより有効である. また, 表8に示すシステムの総合満足度より, 提案システムが満足できるものであったことがわかる.

4. まとめ

本稿では, 感性を用いた異種メディア相互検索システムの評価を行い, システムの有効性を検討した. その結果, ランダムで検索するシステムより有効であることがわかった. しかし, 画像と音楽を合わせようとする心理的傾向も事実であるので, この点についてさらに検討する必要がある. また, 画像によ

表7 人別評価

	system1		system2		system3	
	人数	割合 [%]	人数	割合 [%]	人数	割合 [%]
5.0-4.5	4	20.0	0	0	0	0
4.5-4.0	4	20.0	2	10.0	0	0
4.0-3.5	7	35.0	7	35.0	0	0
3.5-3.0	5	25.0	10	50.0	5	25.0
3.0-2.5	0	0	1	5.0	4	20.0
2.5-2.0	0	0	0	0	7	35.0
2.0-1.5	0	0	0	0	3	15.0
1.5-1.0	0	0	0	0	1	5.0

表8 総合満足度

満足度	system1	system2	system3
	人数	人数	人数
5	8	0	0
4	9	8	0
3	3	10	6
2	0	2	8
1	0	0	6

てはデータベース内に適した音楽が無いのでデータを増やす必要がある. さらに, 文献[10]のように音楽によって画像, 映像にどのような効果を与えたいのかという観点からの検討も考えられる. 以上の課題を改善し, 本システムの更なる性能向上を目指す.

謝辞

本研究の一部は関西大学重点領域研究ならびに関西大学学術研究助成基金の援助のもとで行われた.

文献

- [1] 戸嶋朗, 八村広三郎, “絵画からの画面構成の抽出と検索への応用”, 情処学論, Vol.40, No.3, pp.912-920, Mar. 1999.
- [2] 美濃導彦, 岡崎洋, 坂井利之, “対象物の属性特徴による画像検索法 - 風景画像中の山を例として -”, 情処学論, Vol.32, No.4, pp.513-522, Apr. 1991.
- [3] 串間和彦, 赤間浩樹, 紺谷精一, 山室雅司, “色や形状等の表層的特徴量にもとづく画像内容検索技術”, 情処学論, Vol.40, No.SIG 3, pp.171-184, Feb. 1999.
- [4] 宝珍輝尚, 高田真介, 都司達夫, “感性に基づくマルチメディアの相互アクセスについて”, 情報処理学会データベースシステム, Vol.124, No.3, pp.17-24, May 2001.
- [5] 宝珍輝尚, 都司達夫, “感性に基づくマルチメディアデータの関連付けに関する一考察”, 信学技報, CS2000-113, pp.19-24, Dec. 2000.
- [6] 古賀広昭, 下塩義文, 小山善文, “画像に合った音楽の選定技術”, 信学技報, HCS99-34, pp.25-32, Sep. 1999.
- [7] 林坂享祐, 梶川嘉延, 野村康雄, “異種メディア相互検索システムに関する検討 - 画像に見合う音楽について -”, 信学技報, HIP36-42, pp.33-38, Dec. 2002.
- [8] 池添剛, 梶川嘉延, 野村康雄, “音楽感性空間を用いた感性語による音楽データベース検索システム”, 情処学論, Vol.42, No.12, pp.3201-3212, Dec. 2001.
- [9] 椋木雅之, 田中大典, 池田克夫, “対義語対からなる特徴空間を用いた感性語による画像検索システム”, 情処学論, Vol.42, No.7, pp.1914-1921, Jul. 2001.
- [10] 岩宮眞一郎, “映像作品における視聴覚コミュニケーション”, 信学技報, HIP36-42, pp. 39-46, Dec. 2002.