

H-027

映像コンテンツの構図と時間変動に基づいた情報オーバーレイの一検討 Information Overlay Method Corresponding to the Composition and Temporal Dynamics of the Video Content

高橋 知彦[†] 菅野 勝[†] 酒澤 茂之[†]Tomohiko TAKAHASHI[†] Masaru SUGANO[†] and Shigeyuki SAKAZAWA[†]

1. はじめに

VOD やインターネットストリーミングにおいて、映像表示プレーンを利用し、視聴者へのお知らせや、商品の広告提示を行ったり、メール等のアプリケーションからの入力を表示するといったように、テレビ画面に流れる映像上に、静止画とテキストによる情報を重畳（オーバーレイ）するサービスが検討されている。例えば、画面上の一定領域に、静止画と、テキストによるヘッドラインを重畳し、視聴者が、興味がある情報を見つけた場合、テレビのリモコンを用いて取得する、ティッカーアプリケーション等は、その一例である。

こうした情報重畳においては、コンテンツの魅力が極力損なわれない、画面上の重畳位置・重畳タイミングの決定が重要となる。筆者らはこれまで、画面中の情報量や、重畳後の画面の構図から、最適な重畳位置を自動決定する手法を検討してきた[1]。本稿では、これまでの検討に加え、番組の盛り上がりについても考慮するとともに、提案の手法を実際の映像コンテンツに適用し、評価を行う。

2. 従来手法

映像への情報重畳位置決定に関して、Mei ら[2]は、Saliency Map に基づいた手法を提案している。Mei らの手法は、Motion Attention Analysis に基づいて ROI (Regions-of-Interest) を抽出することで、映像上の注目度の低い箇所を算出し、そこに静止画の広告を重畳するものである。しかしながら、重畳が映像に与える影響という観点では十分な検討がされておらず、広告重畳後に生成される画面のバランスや、番組の盛り上がりは考慮されていなかった。そのため、情報重畳の結果、画面が通常良いとされている構図[3]と大きく異なった構図となり、視聴者が違和感を感じる、あるいは、コンテンツの盛り上がるシーン、即ち映画のクライマックスやスポーツの得点シーンに情報が重畳され、視聴者が集中できなくなってしまうといった問題が考えられる。そこで本稿では、映像コンテンツに対して最適な情報重畳を行うことを目的として、視聴者の視聴体験の変化を最小にするような時間位置及び空間位置に、情報の重畳を行う。

3. 映像に対する情報重畳のアーキテクチャ

本稿で想定する、情報重畳のアーキテクチャを図1に示す。まず、映像配信プラットフォームには、複数のサービスプロバイダや広告主から集められた静止画やテキストが格納されているものとする。他方、映像コンテンツは、本稿の提案する情報重畳決定手法によって解析され、コンテンツに対する最適な情報重畳の時間位置・空間位置が決定されているものとする。ユーザが映像の配信要求を出すと、映像配信プラットフォームは、ユーザやコンテンツの状況に合わせて、静止画やテキストを選択・決定し、映像コンテンツに重畳して、配信する。

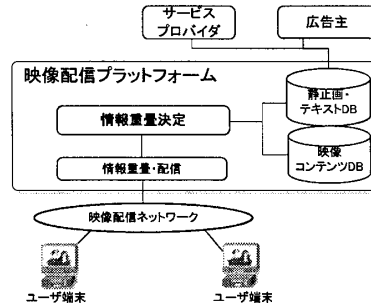


図1: Overlay video advertising のアーキテクチャ

以下、本稿で述べる情報重畳決定手法は、このような、サービスプロバイダから共通的に利用されるプラットフォーム[1]によって利用されることを想定する。

4. 提案手法

本稿では、最適な情報重畳のために必要な要件として、以下の3つを仮定する。

- (1) 映像コンテンツの盛り上がりに対して、重畳された情報が妨げにならないこと
- (2) 重畳によって、多くの情報量を持つ重要な映像要素が隠蔽されないこと
- (3) 重畳した結果、視聴者に提示される映像が不自然とにならないこと

以下、提案手法の具体的な処理について述べる。まず、要件(1)より、番組の盛り上がりの大きい区間を、重畳候補から取り除く。番組の盛り上がりについては、番組の種類によってその特性が異なるが、本稿の検討では、スポーツ番組を対象とする。スポーツ番組においては、番組の盛り上がりは歓声の大きな個所として出現する。そこで、文献[4]の手法を元に、番組中の音声を、(a)会話、(b)音楽、(c)歓声、(d)無音の4種に分類する。さらに、その音量を取得し、歓声が急激に大きくなる時間位置を特定することで、その前後のシーンを、情報重畳の候補から取り除いた。

要件(2)の観点から、動物体上に情報を重畳することは、視聴者から情報の多い領域を隠蔽することになり、好ましくない。そこで、情報重畳の候補領域を求め、それぞれについて、その安定性を算出する。まず、テロップ等、通常のテレビ番組に一般的に重畳される情報の位置に基づき、図2の①～④の4つの領域を、情報重畳の候補領域とする。候補領域それぞれについて、輝度と色差の1秒毎の変化量を算出し、情報を重畳する固定の時間長（本稿では一律10秒とする）において、変化量の総和が少ない領域を、安定領域とする。さらに、要件(2)より、安定領域それぞれに対して、顔検出、テロップ検出、重要物体検出[5]を行い、領域がそれらと重なっている場合は、候補から除外する。これらの処理により残った領域が、情報重畳の候補となる。図3に、映像の再生時間を横軸とした、情報重畳領域と時間の関係の概念図を示す。図中の矢印の区間が、これまでの処理で残った候補領域である。

[†] (株) KDDI 研究所

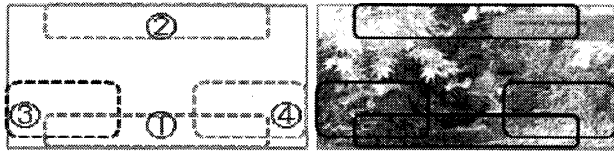


図2：情報を重畳する候補領域(左)と、映像との関係(右)

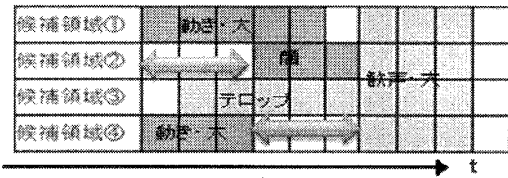


図3：情報重畳候補領域と時間の関係

要件(3)に基づき、取得した各情報重畳候補領域を、情報重畳後の画面の構図に応じて評価を行い、最終的な情報重畳位置・時間を決定する。まず、画面中の構図という観点から、映像本来の注視領域と、情報が重畳される領域の間で適切な距離を保つこととする。また、映像の左右端に情報を重畳する場合は、映像中で良い構図とされる、対角線構図を考慮し、画面中の対角線方向に注視を集める事とする。具体的には、以下の処理を行う。

1. 重畳の候補となる領域を含む代表画像全体から、映像本来が持つ注視領域として、要件(2)で用いた注視領域、即ち、(I)顔、(II)テロップ、(III)情報量の多い領域[5]、(IV)動き ROI の領域をそれぞれ取得し、それぞれの領域の重心 (Distinct Cn) を算出
2. 重畳候補領域の重心と、最も近い位置にある Distinct Cn との間で、重心間の距離 d を算出

この重心間の距離が近い場合、画面中のバランスが悪く、偏った構図になるため、重畳位置の候補の中から、重心間の距離が十分に大きい情報重畳位置を選択する。但し、映像の下端に情報重畳が可能な場合は、三角構図のような安定感のある画面になることが期待できるため、多少重心距離が近い場合でも、下端への重畳は優先的に選択することとした。図4に、構図評価の例を示す。この例では、2つの注視領域と、2つの情報重畳候補領域、③と④がある。各候補領域と注視領域の距離 $d1$ と $d2$ を比較することで、その構図が評価され、より長い距離を持つ領域、即ち領域④が選択される。

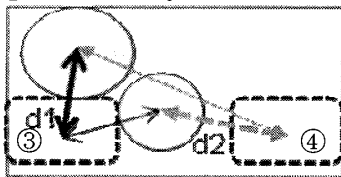


図4：構図評価における重心間の距離の関係

5. 実験

実際のテレビ映像を対象に、情報の重畳を行った。実験は、野球2種、サッカー2種の計4つのスポーツ映像を使用し、それぞれ15~20分の番組を提案手法によって解析した。10秒間の情報重畳を行うとして、情報重畳の候補領域を取得、それぞれに対して構図・盛り上がりに基づく評価を行った。

他方、評価用の正解データとしては、筆者らによる予備実験によって、番組映像全体から、5秒間毎にサムネイルを取得し、各サムネイルに対して、情報を重畳した場合に“視聴の妨げになる”と感じた領域を、任意の数だけ選択し、結果を収集した。提案手法が算出した領域が、被験者

が選択した領域とまったく重複がない場合、重畳は正解と判定した。

表1に、提案手法によって決定された重畳領域を、評価用の正解データと比較した結果を示す。全てのコンテンツにおいて、提案手法によって構図評価値が1位となった領域は、主観評価でも、情報重畳に適切と判別されており、正解率は100%であった。上位10位までは、その正解率は75.0%であった。上位10位までで、重畳が適切にされなかった原因としては、動き ROI や顔領域等、下位特徴量の検出漏れによって、結果として不適切な位置に重畳されたものであり、下位特徴量が正しく取得された場合、構図決定の誤りによって、不適切とされた重畳は無かった。特に、サッカー映像の1つは、遠い距離からゲーム全体を映しているシーンにおいて、正答率が低く、その映像を除いた正解率は83.3%であった。これは、小さく表示されている選手が注視領域として認識されていないことが原因で、動き ROI の閾値調整によって改善できると考えられる。他方、対象を上位20位までに広げると、その正解率は71.7% (サッカー1を除くと66.3%)まで低下した。20位付近の不正解には、情報重畳位置として、テロップのすぐ近くが選択されるなど、不適切な構図となったことが原因の誤りが出現していた。以上から、提案手法によって、映像の構図を評価し、評価の高い領域を選択することが、適切な情報重畳に有効であるといえる。図5に、提案手法によって選ばれた典型的な重畳例を示す。

表1：提案手法による重畳と、人間が許容する領域の相関

| | 提案手法の評価値1位の正解率(%) | 提案手法の評価値上位10位までの正解率(%) | 提案手法の評価値上位20位までの正解率(%) |
|-----------------|-------------------|------------------------|------------------------|
| 全体 | 100.0% | 75.0% | 71.7% |
| 全体からサッカー1を除いた評価 | 100.0% | 83.3% | 66.3% |

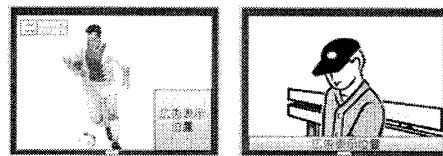


図5：典型的な情報重畳の例(左：サッカーにおけるコーナーキック準備のシーン、右：野球におけるベンチの監督のシーン)

6. 結論

本稿では、映像コンテンツを対象に、画面中の情報量、構図、そして番組の盛り上がりを考慮して、最適な情報重畳の時間位置と空間位置を決定する手法を提案し、その有効性を確認した。本稿ではコンテンツとしてスポーツ映像に限定して評価を行ったが、今後の課題としては、より広範な番組における有効性検証と、精度向上が挙げられる。尚、本発表の一部は、情報通信研究機構からの受託研究“端末プラットフォーム技術に関する研究開発”の開発成果である。

文献

- [1] 高橋、菅野、酒澤、"画面内構図解析を利用した映像コンテンツへの情報オーバーレイ", 映像メ 2010 年次大会
- [2] T. Mei, X-S. Hua, and S.Li, "VideoSense: A Contextual In-Video Advertising System", IEEE Trans. Circuit. Syst. Video Tech., vol. 19, No. 12, Dec. 2009
- [3] R. Thompson, "Grammar of the shot", Focal Press, 1998, p.23
- [4] M. Sugano, Y. Nakajima, H. Yanagihara, and Akio Yoneyama, "Generic Summarization Technology for Consumer Video", PCM 2004, LNCS 3332, pp. 1-8, 2004
- [5] T.Takahashi, M.Sugano, and S.Sakazawa, "Important Object Detection from TV Programs Based on Production Technique Estimation", ICCE 2010