論 文

Technical Papers

閲覧者によるオンラインビデオコンテンツへのアノテーションとその応用

Web-based Video Annotation and its Applications

山本 大介
Daisuke Yamamoto
名古屋大学大学院 情報科学研究科
Graduate School of Information Science, Nagoya University
yamamoto@nagao.nuie.nagoya-u.ac.jp, http://www.nagao.nuie.nagoya-u.ac.jp/members/yamamoto.xml

長尾 確
Katashi Nagao
名古屋大学 エコトピア科学研究機構
EcoTopia Science Institute, Nagoya University
nagao@nuie.nagoya-u.ac.jp, http://www.nagao.nuie.nagoya-u.ac.jp/members/nagao.xml

keywords: video annotation, Web, video retrieval, video simplification, community support system.

Summary

In this paper, we developed a Web-based video annotation system, named iVAS (intelligent Video Annotation Server). Audiences can associate any video content on the Internet with annotations. The system analyzes video content in order to acquire cut/shot information and color histograms. And it also automatically generates a Web page for editing annotations. Then, audiences can create annotation data by two methods. The first one helps the users to create text data such as person/object names, scene descriptions, and comments interactively. The second method facilitates the users associating any video fragments with their subjective impression by just clicking a mouse button. The generated annotation data are accumulated and managed by an XML database connected with iVAS. We also developed some application systems based on annotations such as video retrieval, video simplification, and video-content-based community support. One of the major advantages of our approach is easy integration of hand-coded and automatically-generated (such as color histograms and cut/shot information) annotations. Additionally, since our annotation system is open for public, we must consider some reliability or correctness of annotation data. We also developed an automatic evaluation method of annotation reliability using the users' feedback. In the future, these fundamental technologies will contribute to the formation of new communities centered around video content.

1. はじめに

近年，ハードディスクの大容量化や携帯電話等の簡便な動画記録デバイスの大衆化，さらにはブロードバンド環境の普及に伴い，インターネット上に多数の動画コンテンツが公開されるなど，ネットワークを通してデジタルビデオコンテンツに容易にアクセス可能な環境が整備されつつある．さらには，デジタルビデオカメラの普及や動画編集ソフトウェアの低価格化などにより膨大なビデオコンテンツが氾濫することが予想される．それに伴い，ビデオコンテンツの意味的な検索や要約などを行いたいという欲求は日に日に高まっている．

動画コンテンツの意味的な検索や要約をするためには，そのコンテンツへのメタ情報の付与（アノテーション）が不可欠である．MPEG-7 [MPEG 02] のように動画コンテンツに対するアノテーションの規格が制定されている．また，動画に対する詳細なアノテーションを行う研究は様々なものが行われている [Davis 93, Ricoh 02, Smith 00, Nagao 01, Nagao 02, Nagao 03] が，人的コストが高く，作成に時間がかかる．

本研究では，動画の場合，多くの閲覧者を獲得することが比較的容易であるという点に着目し，その閲覧者からのフィードバックやアノテーションに参加してもらえる環境を整備すれば，より多くのアノテーション情報が集まるのではないかという観点から，一般的な Web ブラウザを用いて，閲覧者による簡単かつ負担の少ない手段で動画に対してアノテーションを行うシステムが有用ではないかと考えた．この方法だと，たとえ一人当たりのアノテーションの量が少なくても，複数の閲覧者のアノテーション結果を融合させることにより，全体として高度なアノテーションとその活用（検索・要約など）が実現できると思われる．そのために，コンテンツを閲覧しつつついでにビデオコンテンツの時間軸に沿って電子掲示板感覚でアノテーションを行うシステムが有用であると考えた．

また，本システムによって得られたアノテーションを

用いた応用例として，ビデオコンテンツの意味的検索および簡約をするシステムを試作した．また，アノテーションの副次的利用として，アノテーション情報をユーザ間で共有することにより，ビデオコンテンツを中心としたコミュニティ形成の手段の一つとして利用できるのではないかとも考えた．それを発展させることによって，新しい広告媒体として利用可能なシステムへの足掛かりとして利用できると考えている．

2. 閲覧者によるビデオアノテーションとその問題

ビデオコンテンツにアノテーションを行う方法には様々なものが考えられるが，筆者は大きく分けて以下の 3 種類があると考えている．

- 自動ビデオアノテーション
- 半自動ビデオアノテーション
- オンラインビデオアノテーション

自動ビデオアノテーションの例としては，音声認識技術や画像認識技術を用いた Informedia [Wactlar 96] や，画像認識技術とオブジェクト指向型状態遷移モデルを組み合わせた研究 [佐藤 96] など様々な研究が行われている．自動ビデオアノテーションは人間の手が介在しないためにアノテーションコストが低いという利点がある一方，コンテンツの種類によって解析手法や解析精度が異なるため一般的なコンテンツに対して適用することが困難だという欠点がある．

半自動ビデオアノテーションの例としては，専用のツールを用いてアノテーションを行う研究がいくつか存在する [Davis 93, Ricoh 02, Smith 00, Nagao 01, Nagao 02, Nagao 03]．人間の手が介在する分アノテーションコストがかかるものの，意味的な情報を付与することができるのは大きな利点である．筆者らも以前，カット検出やオブジェクトトラッキング，音声認識などの技術を用いたアノテーションエディタの試作を行った [山本 02] が，詳細なアノテーションを行うためには，コンテンツの長さの数倍もの時間がかかるのが現状である．また，アノテーション情報をビデオ区間の包含関係に基づいて自動継承する OVID モデル [Oomoto 93] などもアノテーションコストを下げる点において有意義な方法である．

最後に，オンラインビデオアノテーションとは，アノテーションに必要な情報をネットワークを通じてリアルタイムに収集し応用に反映させる手段である．他メディアと放送コンテンツをリンクさせ最新の情報を得る研究 [谷田部 96, 佐藤 95] や，動画像コンテンツに対して電子掲示板感覚で情報を付与する SceneNavi [山田 01] などの研究はこれに当たると考えられるが，アノテーションとしての利用や，応用研究，さらには後述する問題などに対処できていない．

そこで本論文では，オンラインビデオアノテーションとして，閲覧者よりオンラインで情報を収集しアノテーションとして反映させる，閲覧者によるビデオアノテーションシステムを提案する．

一般に，ビデオコンテンツへのアノテーションには多大な人的コストがかかり，各種ビデオ解析処理の高速化や精度の向上を図ってもそれほど改善されない．それは，ビデオコンテンツの深い意味情報付与には人間の高度な判断や解釈を必要とするからである．そこで本研究では，人的資源の不足を補うために，閲覧者が閲覧時に比較的負担の少ないやり方でアノテーションを行う仕組みを提案する．つまり，ビデオコンテンツを解析するという問題を，複数のユーザからのアノテーション情報の収集と解析という問題に帰着させることにより，効率よく動画を解析しようとする試みである．しかも，電子掲示板感覚なインタフェースを採用することにより，閲覧者による自発的なアノテーションが促され，人的なアノテーションコストを最小化する試みである．

ここで，閲覧者によるビデオアノテーションを行う上で解決すべき問題はいくつかあるが，まず，アノテーション情報の量を確保すると同時に情報の質を確保する必要があるという一見相反する問題があげられる．閲覧者による映像コンテンツへのリアルタイム掲示板書き込みの例として，大規模匿名掲示板である 2 ちゃんねる (http://www.2ch.net/) の実況板がある．人気のある TV コンテンツに対しては 1 分間に数十から百以上もの書き込みがあることは珍しくなく，このような匿名掲示板システムならば，アノテーションの量を確保できる可能性が高い．しかしながら，信頼性の高い良質な情報もあるが多くは信頼性の低い情報や日本語として正しくないものであり質は低い．そこで，本システムでは，個々のアノテーションに対してアノテーション信頼度という指標を導入し，アノテーション情報の選別を行うことによって，この問題の解決を図る．

さらに，複数のアノテータのアノテーション情報をいかに統合するかという問題もある．Pradhan らの研究 [Pradhan 97] では，複数のアノテータによるビデオアノテーション間に，矛盾や不完全性が認められた場合に，これらを抽象化する事で解消を図る手法を提案しているが，本システムの場合，アノテーション記述内容の信頼性が低いものが多くあると想定されるため，直接この仕組みを適用することが困難である．この問題にもアノテーション信頼度を用いれば，信頼度付きアノテーション情報を基にして確率的な処理を行うことができ，より現実的な解決法が得られるのではないかと考えている．

また，本システムでは，アノテーションが増えれば，逐次反映され，検索などの応用例の精度向上に結びつき，この点でもオンラインであることの利点が活かせると思われる．

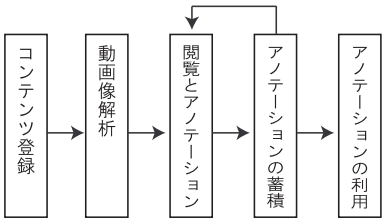


図 1 処理の流れ

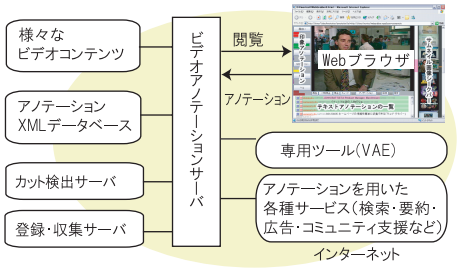


図 2 システム構成

最後に，アノテーションによって得られた情報を用いて，様々な応用例を考察し，アノテーションの有用性を示すことで，従来研究からの差別化が図れると思われる．

3. アノテーションシステムの構築

閲覧者によるアノテーションの有用性と，前章で述べた問題に対する解決法を検証するために，閲覧者によるビデオアノテーションシステムである iVAS(intelligent Video Annotation Server) を構築した．

具体的な処理の流れを図 1 に示す．

3·1 システム構成

本システムの構成図を図 2 に示す．ユーザは，ネットワークからアクセス可能な任意のビデオコンテンツに対して，iVAS を通してアノテーション及び閲覧を行うこととする．また，ユーザは匿名，記名に関わらずアノテーションできる仕組みである．iVAS を通して閲覧したいコンテンツは，登録サーバを用いて明示的に登録する必要があるが，将来的にはコンテンツ収集サーバを用いて自動収集することを考えている．コンテンツを登録すると直ちに，カット検出サーバが連動しカット検出やヒストグラム情報の取得などの自動処理が行われる．ビデオアノテーションサーバによって生成されたページを通してコンテンツを閲覧しつつ，閲覧者がアノテーションを投稿する仕組みである．投稿されたアノテーションはアノテーション XML データベースに蓄積され，6 章で述べる各種のアノテーションを利用したサービスなどで利用される．アノテーションシステムとしては，Proxy Model [Hirotsu 99] に相当する．

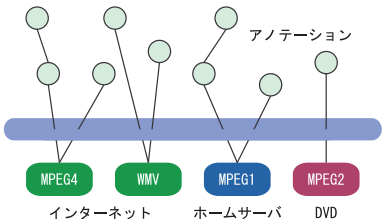


図 3 アノテーションを想定するビデオコンテンツ

3·2 想定するビデオコンテンツ

アノテーションが可能なビデオコンテンツは，PC からアクセスできるデジタルビデオコンテンツである．インターネット上で公開されている Web ビデオコンテンツだけでなく，個人的に HDD ビデオレコーダなどで大量かつ無作為に録り貯めた TV 映像などのホームネットワークに存在するコンテンツ，あるいは DVD などのパッケージメディアコンテンツなどにも適用可能である (図 3)．また，本システムはメタ情報のみを編集・蓄積するものであり，オリジナルコンテンツの編集・コピー・再配信を行うものではないため，著作権的な問題も発生しにくいと考えている．

3·3 コンテンツ登録

ビデオコンテンツは，コンテンツ登録サーバを通してあらかじめ解析を行う．コンテンツ登録時に，テンプレートの選択によるアノテーション作成環境を設定でき，コンテンツに適したアノテーションを作成することができる．

3·4 動画像の解析

Web ブラウザを用いて動画のアノテーションを行う場合，インタラクティブに動画の解析処理をすることは処理速度などの点で好ましくないので，あらかじめ解析を行いたいコンテンツに対して前処理を行う必要がある．そのために，あらかじめカット検出を行い，動画からカットの時刻とサムネイル画像をサーバ上に保存するプログラムとしてカット検出サーバを用意した．

カット検出のアルゴリズムは，現在のフレーム間差分と直前のフレーム間の比較を行い，ある閾値を超えた時点でカットとした．また，フレーム間差分算出手法には，分割 χ 二乗検定法 [長坂 92] を採用した．また，フレーム間差分の変動が大きい区間（つまり動きの激しい区間）はひとつのカットと認識するなどの工夫をしている．

カット検出サーバの対象とするコンテンツは，カットが多く存在するコンテンツであり，サッカー中継や個人で撮影したホームビデオや講演ビデオなどカットの少ないコンテンツにはカット検出が適用できない．これらのコンテンツは，一定時間ごとに区切り，形式的なカットとし，アノテーションを行うこととする．

また，カット検出の過程で得られる色ヒストグラム情報も XML データベースに蓄積する．

表 1 カット検出の精度．検出カット数は本システムでカットと認識した数，過検出は本来カットではない部分をカットとして認識した部分，未検出は本来カットである部分を検出できたかった数をいう．ニュースの精度が悪いのは，花火大会の映像で誤検出が多かったためである．

	ジャンル	検出カット数	過検出	未検出
a	ニュース	56	8	8
b	ドラマ	31	0	2
c	バラエティ	72	1	3
d	料理番組	29	0	3

今回の実験で用いたコンテンツ（詳しくは 7 章で述べる）に対するカット検出の精度を表 3 に示す．全体の適合率は 89.3 %，再現率は 91.7 %であった．

4. 閲覧者によるオンラインビデオコンテンツへのアノテーション

閲覧者は，iVAS のアノテーション編集ページを用いて，テキスト入力を主としたアノテーション（テキストアノテーション）とマウスクリックを主にしたアノテーション（印象アノテーション）及びテキストアノテーションに対する○×評価（評価アノテーション）の 3 つを行うことができる．

また，アノテーションモデルとしては，代数ビデオ [Weiss 94] の stratified model に相当する．

4·1 アノテーション編集ページ

アノテーション編集ページのブラウザは図 4 のようになる．画面左に印象アノテーションインタフェース，中央部上部に動画閲覧画面，画面中央下部にテキストアノテーションの一覧，右側にサムネイル画像を用いたスクロール可能なシークバーを配置した．なお，カット検出直後のフレーム画像は乱れていることが多いため，カットとして検出したフレームから 20 フレーム後をサムネイルとして選択している．

このシークバーは，マウスのスクロールボタンによってシームレスにシーク可能なバーであり，アノテーションを行う時に頻繁に繰り返されるビデオのシークを直感的に支援する．基本的には閲覧することが主目的であるので，なるべく動画閲覧画面を大きくとる構成にしている．

また，テキストアノテーションの一覧は，現在のカットに関連する情報を時間軸に応じて表示している．また，後に述べる重要度に応じて，重要度の高い情報が上位に来るようにソートして表示することにより，多数の情報を効率よく表示している．

4·2 テキストアノテーション

テキストアノテーションは，ビデオ内のシーンやオブジェクトに対して，テキストで情報を付加するものである．

図 4 アノテーション編集ページ

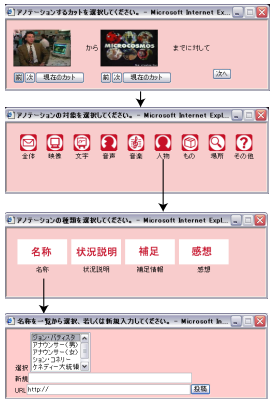


図 5 テキストアノテーションの例．「この人物の名称はジョン・バティスタだ」という内容を投稿している．

まず，ユーザはアノテーションしたい場面で動画上のアノテーションしたいオブジェクトのある部分をクリックする．このクリックした位置を取得することにより，オブジェクトの画面上のおおよその位置が取得可能である．次に，カット検出サーバ等であらかじめ検出されているカット単位でアノテーションしたい時間範囲を選択する．さらに，全てのアノテーションには，後の検索や要約等の機械処理をしやすくするために，コメントの対象（全体・映像・キャプション・音声・音楽・登場人物・オブジェクト・場所など），種類（名前・状況説明・補足情報・感想など）を選択肢をわかりやすく表示し，容易に選択できるようにした．さらに，他の閲覧者によって入力されたテキストアノテーションに対し，閲覧者が評価する仕組み（○・×のボタンを押す）を用意し，閲覧者によって個々のアノテーションに対する評価を行うことができるように配慮した．コメントの対象と種類のカテゴリは適宜追加・編集可能であるが，どのようなカテゴリを用意するかは今後の課題とする．具体的な例を図 5 に示す．結果は XML データベースに格納される．

まとめると表 2 のような情報を取得することができる．

4·3 印象アノテーション

印象アノテーションとは，ビデオコンテンツの雰囲気や閲覧者の主観的印象，例えば，面白い・緊迫・悲しい

表 2 取得可能なアノテーション情報とその取得方法

アノテーション情報	取得手段	作成条件
アノテーション固有 ID	投稿時に生成	自動
個人情報	Cookie で入力	自動
オブジェクトの位置	動画上をクリック	暗黙的
時間範囲	カット単位で指定	必須
コメントの対象と種類	対話的に選択	必須
コメント	テキスト入力	必須
名称	選択または記述	必須
評価	○・×ボタン	任意

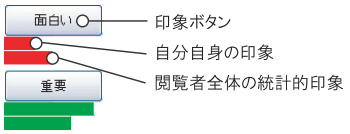


図 6 印象アノテーションのインタフェース．印象ボタンの下に 2 段の棒グラフがあり，ユーザの現在の印象情報の表示と，閲覧者全員の統計的な印象情報を表示する．

などをマウスクリックで入力できる仕組みである．より印象深いシーンでは印象ボタンの連打度合いによって印象の強弱を表現できる．

印象アノテーションの対象となる各印象を $I_1 I_2 ... I_n$ とする．その時，クリックした時間を中心にして，正規分布 $N(\mu, \sigma^2)$ で印象情報をつけるとすると，各印象 I_k は以下の式でパラメータを付与する．

$$I_k(t) = \sum_{\text{印象 } I_k \text{の選択の集合}} N(t_i, m) \qquad (1)$$

ただし t_i は印象アノテーションを行ったメディア時間である．m は定数であり，ボタンを押した時の前後の時間にもアノテーションの効果を与える．

また，自分のアノテーション結果だけでなく，閲覧者全体のアノテーションの結果も棒グラフによって表示している (図 6)．ボタンの数は最大 6 個まで可能であり，これは，登録サーバで指定可能である．どのようなボタンが有効であるか，また何個必要かといった検証はコンテンツの種類に依存する事から今後の課題である．

5. アノテーション信頼度

不特定多数のユーザによるアノテーションを扱うとどうしても，信頼性の低い情報が混在する可能性がある．そのため，各々のアノテーションに対する信頼度を計算し，情報の選別を行う必要がある．信頼度の計算方法として,「信頼できる情報を多数入力した人の情報ほど信頼できる」という原則に基づいて次の方法で計算している．

まず，A_k に対する単純評価 e_k を以下のように求める．おのおののアノテーションに○ (good) の評価をした人の数 g_k と× (bad) の評価をした人の数 b_k とする．また，選択項目に矛盾がないか，日本語として正しい記述をしているかどうかということにより機械的に決まる評価を c_k とする．ただし，c_k は $-1 < c_k < 1$ となり，c_k の値が大きいほど良いアノテーションと機械的に判断されたこととする．なお，自然言語文によるアノテーションの場合，後の検索や要約などの応用において形態素解析が重要となる場合が多い．そこで，形態素解析を行い，文全体の形態素数のうち未知語の含まれる割合（未知語率）や，文の長さ，構文解析の結果等を総合的に評価することにより c_k を求める必要がある．

これにより，単純評価 e_k は以下の式になる．

$$e_k = s \cdot \frac{g_k - a \cdot b_k}{g_k + a \cdot b_k} + t \cdot c_k \qquad (2)$$

s は閲覧者による評価の割合，t は機械による評価の割合であり，$s + t = 1$ とし，機械的な評価の精度に依存して t の値を大きくする．

また，a は○評価と×評価の割合を補正する係数であり，すべてのアノテータが行ったすべてのアノテーションに対する評価の総数を g_{all}, b_{all} とすると，

$$a = \frac{g_{all}}{b_{all}} \qquad (3)$$

と表す．e_k は，$-1 < e_k < 1$ の値をとる．

(2) は機械的な評価と人間の評価を組み合わせた直感的な式であるが，このままでは，アノテーションを行う個人（アノテータと呼ぶ）の信頼性が考慮されていない．

そこで，アノテータに対する信頼度 p を求める．これは今までアノテータが行った全てのアノテーションに対する○ (good) 評価数を G，× (bad) 評価数を B とすると，アノテータ信頼度 p は，

$$p = d(G + B)\frac{G - a \cdot B}{G + a \cdot B} \qquad (4)$$

により求める．ここで，$d(x)$ はサンプル数が少ない場合に評価値を低く抑える関数であり

$$d(x) = 1 - \exp(-\tau \cdot x) \qquad (5)$$

とする．ここで，τ はどの程度評価値を抑えるかを決める定数であり $\tau > 0$ である．

また，動画像の時間軸にそったアノテーションをリアルタイムに閲覧者が評価する場合，刻一刻とアノテーション情報が変化するために，閲覧者が誤った○×評価をする場合も多く，閲覧者による評価が集まっていない場合にはアノテータ信頼度を重視し，閲覧者による評価が集まっている場合には閲覧者による評価を重視する必要がある．そこで，アノテーションに対する信頼度 r_k を以下のようにする．

$$r_k = (1 - d(g_k + b_k)) \cdot p + d(g_k + b_k) \cdot e_k \qquad (6)$$

これにより，アノテーション信頼度 r_k を求めることができる．r_k は，$-1 < r_k < 1$ の値をとり，値が大きいほど相対的に信頼できるコンテンツであると言える．

なお，匿名の書き込みの場合は，アノテータ信頼度は最低の $p = -1$ とする．アノテータに対する信頼度を計算する意義は，機械的にアノテーションを評価するのは難しいこと，ユーザ評価が集まっていない段階ではその情報の信頼性が不明なこと，さらに，信頼性の低い人の大量書込みを防ぐこと（いわゆるアラシ対策），ユーザに信頼性を公開し，信頼される情報を入力するように暗黙的に誘導するところにあると思われる．

6. アノテーションの応用

本システムのアノテーション情報の有効性を示す例として，検索・簡約・コミュニティ支援に関するアノテーション応用システムを作成した．

6·1 自然言語による Web 検索

本研究で得られたアノテーション情報の有用性を確認する例としてビデオコンテンツの意味内容に即した検索を行う．カットに対する色ヒストグラム情報やテキストアノテーション情報を元に検索を行う (図 7)．

基本的な検索アルゴリズムとしては，筆者らが以前の研究 [山本 02] で提案した，自然言語による意味的ビデオ検索の手法を改良した．具体的には，検索文，テキストアノテーション文，共に茶筅 [奈良 03] を用いて形態素解析を行い動詞・形容詞・名詞・未知語を取り出し，それぞれの単語の基本形を基にしてコサイン距離（マッチする単語が多いほど点数を加算）を求める．そして，テキストアノテーションを適用するそれぞれのショットに対してその点数を加算する．さらに今回の手法では，印象アノテーション情報とアノテータ信頼度の情報を用いている．ショットに対してテキストアノテーションの情報を加算する時，アノテータ信頼度と，映像の状況を直接的に表現しているであろうと思われる度合い，つまり，アノテーションの種類が名称 > 状況説明 > 補足説明 > 感想の順に比例した点数を加点している．

さらに，各ショットに対して，印象アノテーションの値が大きいショットほど重要度を上げている．

このようにして加点された得点に基づいて表示順序を決定し，検索結果をユーザに提示する．検索結果例を図 8 に示す．

6·2 アノテーションによるビデオ要約への応用

アノテーションを用いた要約システムは，音声トランスクリプトに付与された言語構造から重要度を計算してビデオ要約を行う長尾ら [Nagao 01, Nagao 02, Nagao 03] の研究などいくつか存在する．しかしながら，閲覧者の盛り上がりや視聴者情報を用いていないために，的確な要約を行っているとはいえない．そこで，もし本研究のシステムで得られたアノテーション情報を適用すれば，視聴者情報を活かしたより的確な要約ができると考えられる．

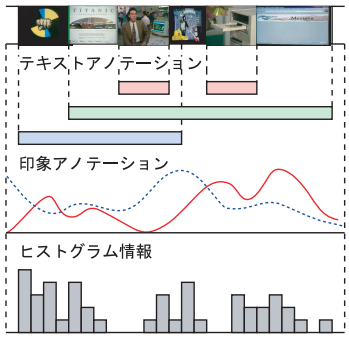


図 7 検索に用いるアノテーション情報

図 8 Web ビデオ検索結果

具体的には，それぞれのカットに対して，印象アノテーションの各アノテーションの印象度の合計値の時間平均を正規化したものと，そのカットに関連するテキストアノテーションの数を足し合わせたものを，そのカットの閲覧者による重要度と考えて，要約をする場合のシーンの重要度として加算すればよいと考えられる．

本システムでは，「盛り上がっているシーンほど重要」という簡単な規則により上記の方式による閲覧者によるアノテーション情報のみを利用してビデオコンテンツから重要シーンを抽出して時間短縮を行うシステムを試作した．具体的には，上記の方法によって求めた重要度を用いて，指定した簡約時間を超えるまで，重要度の高いカットから順に選択していった．本来ならば，ストーリのつじつまが合うように要約をする必要があるが，今回はそこまで行っておらず，簡約，すなわちコンテンツの時間的縮退に留めた．簡約結果を図 9 に示す．

6·3 アノテーションによるコミュニティ支援

アノテーションによって副次的に得られる個人情報を用いた応用として，コミュニティ支援が考えられる．

既存のコミュニティ支援やコンテンツ推薦システムは，コンテンツを見たか見ていないか，あるいはコンテンツを購入したか，していないかによって形成されるものが

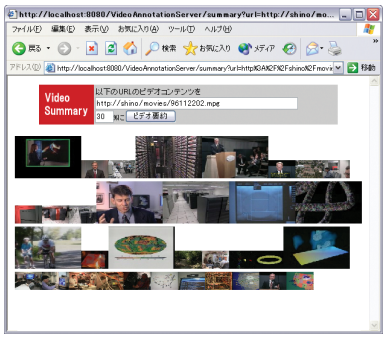


図 9 ビデオ簡約の例（小さいサムネイル画像の部分のカットが省略される）

中心である．コンテンツの意味的内容に応じたコミュニティ形成支援やコンテンツ推薦を実現している仕組みは少ない．しかしながら，コンテンツの内容を加味しなければ正確なシステムの実現が難しい例が多い．例えば，サッカー番組でサッカーの試合が好きな人たちと特定のサッカー選手のみが好きな人たちとは本質的に違うコミュニティであるし，また，だれもが閲覧しているような有名な映画では単に見たというだけでは情報量が小さく，アクション部分が好きなのか，恋愛シーンが好きなのかによって属するコミュニティが異なる．そこで，印象アノテーション情報を用いて，人と人との興味やものの感じ方の近さの一つの指標となる，印象距離を求める方法を提案し，コミュニティ発見やコンテンツ推薦のための一つの指標となれば良いと考える．

まず，あるコンテンツ C をみた人 P_j と P_k に関する印象距離 $D_C(P_j, P_k)$ を印象アノテーション情報を用いて，以下の式で定義する．

$$Dc(P_j, P_k) = \int (I_{pj} - I_{nj}) \cdot (I_{pk} - I_{nk}) dt \quad (7)$$

ここで，I_{pj} は式 (1) で計算される P_j のポジティブな印象アノテーション情報であり，I_{nj} はネガティブな印象アノテーション情報，I_{pk} と I_{nk} はそれぞれ，P_k のポジティブ印象アノテーション情報とネガティブ印象アノテーション情報を示す．

なお，印象アノテーション情報はマウスクリックの頻度に対する個人差が少なくなるように最大値 1，最小値 0 で正規化を行う必要がある．

7. 実験と考察

iVAS システムを利用したアノテーション信頼度と印象アノテーションの評価を行うために以下の実験を行った．

映像処理評価用映像データベース [馬場口 02] から表 3 で示す 5 分程度に編集した評価用映像コンテンツ a,b,c,d の 4 つを用いて，大学生男女 30 人に対してアノテーション及び評価に関する実験を行った．それぞれ，ニュース，ドラマ，バラエティ，料理番組である．印象ボタンは，ポジティブなボタンとして「楽しい」「おいしそう」「重要」，ネガティブなボタンとして「嫌悪」「悲しい」を用意した．なお，1 コンテンツにつき，1 名あたり平均 10 分程度の時間をかけてアノテーションを行っている．得られたアノテーションは表 4 のとおりである．

表 3 評価用映像コンテンツ．利用メディア時間でカット編集した映像を用いている．

	コンテンツ名	利用メディア時間	カット
a	ニュース 19	5 分 33 秒〜10 分 56 秒	56
b	ピエロの涙	0 分 30 秒〜5 分 30 秒	31
c	女王様のブランチ	0 分 00 秒〜6 分 12 秒	72
d	料理番組	0 分 20 秒〜3 分 50 秒	32

表 4 得られたアノテーション情報の数．テキストはテキストアノテーションの書き込み数，印象は印象アノテーションのマウスクリック数，○評価，×評価はそれぞれの評価ボタンを押した数．

	評価人数	テキスト	印象	○評価	×評価
A	30 人	174	2153	239	68
B	30 人	160	1225	234	70
C	30 人	222	2132	334	61
D	30 人	97	1224	159	44

テキストアノテーションに限り，なるべく正確かつ有用な情報を真面目に書き込んでもらうグループ A，半分普通に半分不真面目な情報を書き込んでもらうグループ B，不真面目な情報を積極的に書き込んでもらうグループ C，自由に書き込んでもらうグループ D，テキストアノテーションを行わないグループ E の 5 つのグループに分けて実験を行った．ここでいう，不真面目とは，全く関連のない情報や間違った情報，あるいは記述した内容自体は妥当であるが，表現に不備があることをいう．

また，この実験を行う前に事前アンケートとして，NHK ONLINE MEMBERS(https://members.nhk.or.jp/) の中にある好きな番組ジャンル設定 92 項目に対して，好き・普通・嫌いの 3 段階評価を行った．これは，印象距離に関する考察のために用いた．

7·1 アノテーション信頼度に関する実験と考察

アノテーション信頼度はアノテータ信頼度と閲覧者による○×評価を元にして求める．ここでの○×評価は一般的なものであり，特にバイアスがかかっているわけではない．一方，アノテータ信頼度に関しては，よく考察する必要がある．そこで，アノテータ信頼度の妥当性を考察するために，真面目グループ (A)，半分不真面目グループ (B)，不真面目グループ (C)，自由グループ (D) のアノテータそれぞれに関するアノテータ信頼度を式 (4) より求め 7·1 節に図示した．なお今回の実験では，式 (2) における機械的な評価の精度による誤差を無くす為に $s = 1, t = 0$ とした．また，○の評価が×の評価に比べて 4 倍以上多く，式 (3) より $a = 4$ とした．また，閲

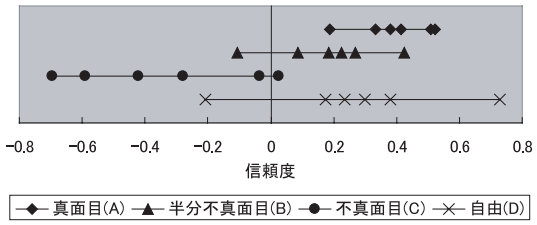


図 10 アノテータ信頼度．数値が大きいほど信頼できるアノテータである．

覧者による評価数が 4 個で式 (5) の値がおよそ 0.5 になる $\tau = 0.2$ とした．

7·1 節から分かるように，アノテータ信頼度の値は A グループ $>$ B グループ $>$ C グループ の傾向があり，これは直感と一致する．信頼度にばらつきが生じる理由は，アノテータによって不真面目さの基準が異なるため，不真面目な書き込みであっても，それは映像から想起される情報であるため，内容によっては閲覧者がその書き込みを容認し×の評価をしない場合があるためである．

また，自由に書き込んだグループのアノテータ信頼度のばらつきが大きいため，アノテータ信頼度を求める意義があると考えられる．

この結果得られたアノテータ信頼度を元にして，D グループ 6 人，153 個のテキストアノテーションに対する式 (6) によって求めたアノテーション信頼度の値と，式 (2) の単純評価を求め妥当性を比較した．まず，筆者が主観で全てのテキストアノテーションを良い・悪いの二つに分類した．分類基準は，シーンの内容を的確に表現していない，あるいは，正しい日本語の記述ではないものを悪いとし，それ以外は良いとした．悪い分類をしたものは 23 個，良い分類をしたものは 130 個あった．このうち単純評価で明らかに間違っているもの (悪い分類なら $r_k > 0.4$，良い分類なら $r_k < -0.4$) は 17 個であるのに対し，アノテーション信頼度では 3 個と，アノテータ信頼度を考慮しない場合に比べて，間違った評価をした信頼度が減少しており，その点で改善されているといえる．ただし，今回の予備実験では，テキストアノテーションの数が 153 個なのに対して，○×評価の数が 289 個と少ないため，正確な評価は今後の課題としたい．

7·2 印象距離に基づくコミュニティの視覚化に関する実験

女王様のブランチ六本木編に対して行った印象アノテーション情報 26 人分を式 (7) を用いてそれぞれの印象距離を求め，ばねモデルを用いて図示したものを図 11 に示す．一つの円が一つのアノテータを意味し，距離が近いアノテータほど今回のコンテンツに関しては興味が近かったことを意味する．女王様のブランチ六本木編は，アメリカ風カフェ，和風料理店，バー，ファッション，カラオケ，映画館，喫茶店，神社関連のお店や名所を順に 1 分程度づつ紹介していく番組である．事前アンケートにより，ファッション番組に興味があると分かっている人について，図 11 に示すように，印象距離を見てみると比較的近い距離に密集する傾向があり，一つのコミュニティを形成していると考えられる．

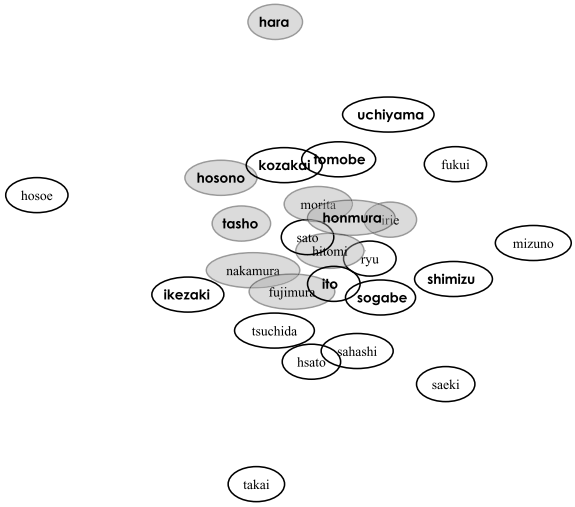


図 11 印象距離．一つの円が一人のアノテータを意味し，距離が近いアノテータほど興味が近い．背景が灰色の円はアンケートでファッション番組に興味があると答えた人，太字の名前は歴史・紀行に関する番組に興味がある人を示す．

図 11 のように，歴史・紀行に興味がある人も印象距離が近い場合があるが，ファッションに興味がある人に比べて密集度が小さく，コミュニティを形成しているとは考えにくい．女王様のブランチのように複数の話題があるコンテンツの場合，人によって興味のもつ話題が複数に分散している場合が多く，そのために式 (7) の計算手法では全体的に印象の薄い話題が印象の強い話題でかき消されてしまう可能性があるからである．明確なコミュニティを視覚化するにはより多くの閲覧者を必要とするか，あるいは，印象距離を話題ごとに時間区間で区切って個別に考える必要があると考えられる．

7·3 アンケートによる評価

本システムは，より多くのユーザに使ってもらえてこそ意義があるシステムである．そこで，システム全体の使いやすさに関するアンケートを行った．以下にその結果を示す．アンケートは，評価実験をしてもらった後にやってもらい，5 段階評価で集計をした．

母集団が大学生ということもあるが，iVAS の使いやすさを示す「取っ付き易さ」の項目で多くの人が普通よりも良い項目をつけており，本システムのインタフェースはなかなか評判のよいものであった．また，iVAS を使ってアノテーションをしたいかという問いに対しても同様に高評価であり，本システムが比較的多くのユーザに使ってもらえる可能性を示唆する結果となった．

表 5 5 段階評価によるアンケート結果．数字が大きいほど良い．なお，テキストアノテーション，印象アノテーションはそれぞれのインタフェースの使いやすさについてのアンケートである．また，テキストアノテーションを行っていない 6 人に関してはテキストアノテーションの評価をしていない．

	1	2	3	4	5	平均
テキストアノテーション	0	3	7	12	2	3.50
印象アノテーション	0	3	8	11	8	3.80
正確に付与出来たか	0	2	11	16	1	3.53
取っ付き易さ	0	2	7	10	11	4.00
iVAS を使いたいか	1	1	11	11	6	3.67

8. お わ り に

本研究では，オンラインビデオコンテンツに対して，不特定多数の閲覧者による，Web ブラウザを用いたビデオアノテーションシステムを構築した．不特定多数のアノテーションで問題となるアノテーションの信頼度の計算方法を提案し，その手法の妥当性を確認した．また，アノテーション情報のいくつかの応用例を示し，評価実験によりその情報の有用性を確認した．

また，今回のアノテーションの応用例は，得られた結果を用いて様々な応用が可能だということを例示したに過ぎない．実際には，要約・コンテンツ推薦・流通広告・コミュニティ形成など，さらに深く多様な応用例が考えられ，極めて拡張性が高いと考えている．

なお，本研究で作成したシステムが以下のページで公開されているのでぜひ参照して頂きたい．

http://www.nagao.nuie.nagoya-u.ac.jp/ivas/

◇ 参 考 文 献 ◇

[Davis 93] Davis, M.: An Iconic Visual Language for Video Annotation., in *Proceedings of IEEE Symposium on Visual Language*, pp. 196–202 (1993)

[Hirotsu 99] Hirotsu, T., Takada, T., Aoyagi, S., Sato, K., and Sugawara, T.: Cmew/U-a multimedia Web annotation sharing system, in *TENCON 99. Proceedings of the IEEE Region 10 Conference*, Vol. 1, pp. 356–359 (1999)

[MPEG 02] MPEG, : *MPEG-7*, MPEG-7 Consortium, http://www.mp7c.org/ (2002)

[Nagao 01] Nagao, K., Shirai, Y., and Squire, K.: Semantic Annotation and Transcoding: Making Web Content More Accessible, *IEEE MultiMedia*, Vol. 8, No. 2, pp. 69–81 (2001)

[Nagao 02] Nagao, K., Ohira, S., and Yoneoka, M.: Annotation-Based Multimedia Summarization and Translation, in *Proceedings of the Nineteenth International Conference on Computational Linguistics (COLING-02)*, pp. 702–708 (2002)

[Nagao 03] Nagao, K.: *Digital Content Annotation and Transcoding*, Artech House Publishers, London (2003)

[Oomoto 93] Oomoto, E. and Tanaka, K.: OVID: Design and Implementation of a Video-Object Database System, *IEEE Transactions on Knowledge and Data Engineering*, Vol. 5, pp. 629–643 (1993)

[Pradhan 97] Pradhan, S., Tajima, K., and Tanaka, K.: Querying Video Databases based on Description Substantiality and Approximations, in *Proceeding of the IPSJ International Symposium on Information Systems and Technologies for Network Society, World Scientific*, pp. 183–190 (1997)

[Ricoh 02] Ricoh, : *Movie Tool*, http://www.ricoh.co.jp/src/multimedia/MovieTool/ (2002)

[Smith 00] Smith, J. R. and Lugeon, B.: A Visual Annotation Tool for Multimedia Content Description, in *Proceedings of SPIE Photonics East, Internet Multimedia Management Systems* (2000)

[Wactlar 96] Wactlar, D., Kanade, T., Smith, A., and Stevens, M.: Intelligent Access to Digital Video: Informedia Project, *IEEE Computer*, Vol. 29, No. 5, pp. 140–151 (1996)

[Weiss 94] Weiss, R., Duda, A., and Gifford, D. K.: Content-Based Access to Algebraic Video, in *Proceedings of IEEE First International Conference on Multimedia Computing and Systems, Boston, MA* (1994)

[佐藤 95] 佐藤 隆, 坂内正夫：ライブハイパーメディアに基づく放送映像のアクセス, 情報処理学会第 51 回（平成 7 年度後期）全国大会講演論文集, 第 3 巻, pp. 301–302 (1995)

[佐藤 96] 佐藤 隆, 坂内正夫：ライブハイパメディアにおける映像情報の獲得, 電子情報通信学会論文誌（D-II）, 第 J79-D-II 巻, pp. 559–567 (1996)

[山田 01] 山田 一穂, 宮川 和, 森本 正志, 児島 治彦：映像の構造情報を活用した視聴者間コミュニケーション方法の提案, 情報処理学会, グループウェアとネットワークサービス, 第 43 巻 (2001)

[山本 02] 山本 大介, 長尾 確：半自動ビデオアノテーションとそれに基づく意味的ビデオ検索, 情報処理学会第 65 回全国大会, No. 3, pp. 95–96 (2002)

[谷田部 96] 谷田部 智之, 佐藤 隆, 坂内 正夫：放送間のリアルタイムリンクを可能とするアドバンスト TV の構想, 電子情報通信学会情報・システムソサイエティ大会 D-279 (1996)

[長坂 92] 長坂 晃朗, 田中 譲：カラービデオ映像における自動索引付け法と物体探索法, Vol. 33, No. 4, pp. 543–550 (1992)

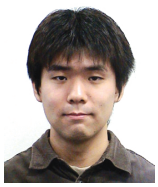
[奈良 03] 奈良先端科学技術大学院大学自然言語処理学講座：形態素解析システム 茶筌, http://chasen.aist-nara.ac.jp/ (2003)

[馬場口 02] 馬場口 登, 栄藤 稔, 佐藤 真一, 安達 淳, 阿久津 明人, 有木 康雄, 越後 富夫, 柴田 正啓, 全 炳東, 中村 裕一, 美濃 導彦, 松山 隆司：映像処理評価用映像データベースについて, 電子情報通信学会技術研究報告 PRMU2002-30 June (2002)

〔担当委員：武田英明〕

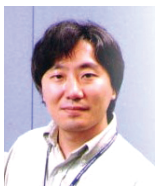
2004 年 4 月 2 日　受理

著 者 紹 介



山本　大介（学生会員）

2003 年 名古屋大学工学部電気電子・情報工学科卒業, 2003 年より 現在 名古屋大学大学院情報科学研究科メディア科学専攻 修士課程.



長尾　確（正会員）

1987 年 東京工業大学 総合理工学研究科 システム科学専攻 修士課程修了, 1987-1991 年 日本アイ・ビー・エム株式会社 東京基礎研究所, 1991-1999 年 株式会社ソニーコンピュータサイエンス研究所, 1994 年 東京工業大学 理工学研究科 情報工学専攻 論文博士, 1996-1997 年 米国イリノイ大学アーバナ・シャンペーン校 ベックマン研究所, 1999-2001 年 日本アイ・ビー・エム株式会社 東京基礎研究所, 2001-2002 年 名古屋大学 工学研究科 情報工学専攻 助教授, 2002-2004 年 名古屋大学 情報メディア教育センター 教授, 2004 年より 現在 名古屋大学 エコトピア科学研究機構 教授.