I-004

# 閲覧者によるオンラインビデオアノテーションとその応用
# Web-based Video Annotation and its Applications

山本 大介 † 長尾 確 ‡
Daisuke Yamamoto Katashi Nagao

## 1. はじめに

近年、ハードディスクの大容量低価格化に登場に伴い、ハードディスクビデオレコーダや Web ビデオコンテンツなどのデジタルビデオコンテンツが普及しつつある。それに伴い、ビデオコンテンツを要約したい・検索したいという要求が高まっているが、意味的な検索・要約をするためには、MPEG-7 などに代表される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。しかしながら、動画に対する詳細なアノテーションを行う研究はいくつかあるが、[1, 2, 3, 4]、人的コストが高く作成に時間がかかる。そこで、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ないかと考えた。この方式だと、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。

また、本システムによって得られたアノテーションの応用例として、自然言語によるビデオコンテンツの検索および簡約をするシステムを試作した。

## 2. iVAS : intelligent Video Annotation Server

本研究ではユーザがコンテンツを閲覧しつつ、アノテーションできる環境を作成し、閲覧者にアノテーションを促すウェブシステムとして iVAS(intelligent Video Annotation Server) を構築した。

### 2.1 システム構成

本システムの構成図を図 1 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、ビデオアノテーションサーバを通してアノテーション及び閲覧を行う事とする。アノテーションを行いたいコンテンツは、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解析が行われる。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、5 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。

†名古屋大学 大学院情報科学研究科
‡名古屋大学 情報メディア教育センター

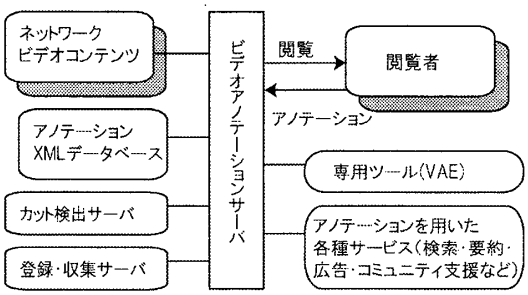


図 1: システム構成

### 2.2 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り以下のものが上げられる。

- 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
- 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 映像コンテンツ
- DVD などのパッケージメディアコンテンツ

DVD などのメディアコンテンツはその ID を、ネットワーク上の Web コンテンツは URI(Universal Resource Identifiers) を、ホームサーバ上の TV コンテンツは EPG(Electronic Program Guide) 情報をキーとして、コンテンツを一意に識別することを考えている。

ここで強調したいのは、元のコンテンツを一切改変しないこと、またあくまでも個人で所有されているコンテンツや閲覧する権利のあるオンラインビデオコンテンツを対象としている点である。

### 2.3 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事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も作成した。サーバプログラムとして動作するために、Java 等との連携や複数同時リクエストに対応している。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。

## 3. 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


図 2: アノテーション編集ページ

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）によってアノテーション情報を投稿する事ができる。

### 3.1 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 2 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。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。

### 3.2 対話式テキストアノテーション

対話式テキストアノテーションは、任意の連続するカットに対して、対話式インタフェースによりテキストでアノテーションする方式である。

動画上のアノテーションしたいオブジェクトをクリックし、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）を順次対話形式で選択できるようにした。さらに個々の書き込みに対し、閲覧者が評価する仕組み（○・×のボタンを押す）を用意した。

さらに、アノテーションを行ったユーザ名、E-mail などの個人情報、コンテンツ ID、日時などが自動的に記録されインターネット上の XML データベースに投稿される。

投稿できるアノテーションを表 1 にまとめる。

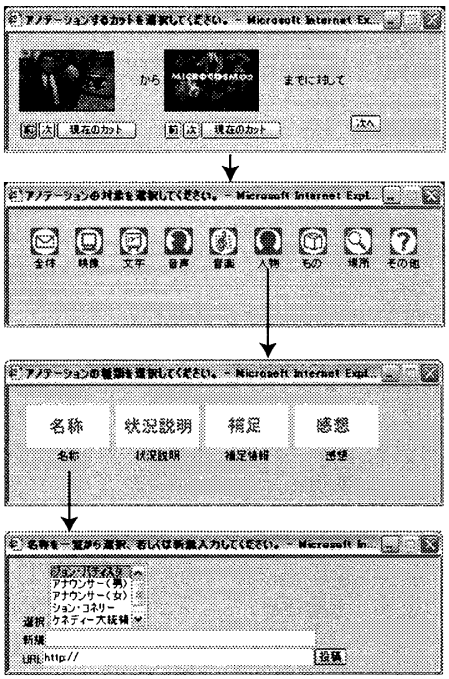


図 3: 対話的なテキストアノテーションの例

表 1: 得られるアノテーションとその取得方法

| アノテーション | 取得手段 | |
|---|---|---|
| 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 |
| 本文 | テキスト入力 | 必須 |
| 名称 | 選択または記述 | 必須 |
| 関連 URL | テキスト入力 | 任意 |
| 評価 | ○・×ボタン | 任意 |

「画面上に登場している人物の名称はジョン・バティスタだ」という内容を記述する例を図 3 に示す。

### 3.3 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・悲しいなどをマウスクリックでアノテーション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を行う各印象を $I_1 I_2 ... I_n$ 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。

$$I_k(t) = \sum_{i=\text{印象 } I_k \text{のアノテーション}} N(t_i, m)$$

ただし $t_i$ は $I_k$ に関する $i$ 番目の印象アノテーションをし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。

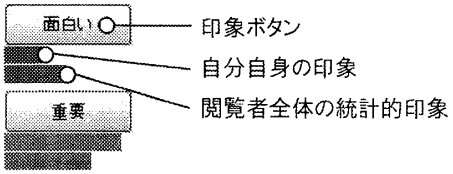


図 4: 印象アノテーション

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 2 左, 図 4)。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ことから今後の課題である。

## 4. 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存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、「信頼できる情報をたくさん書き込んだ人の情報ほど信頼できる」という原則で次の方法で計算している。

あるアノテーション $A_k$ に対する信頼度（アノテーション信頼度）は以下のように求める。まず、$A_k$ に対する単純評価 $e_k$ を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 $b_k$、さらに項目がきちんと選択されている、形態素解析が良好などの機械的な評価を $c_k$ を用いて、以下の式で求める。

$$e_k = s \cdot d(g_k + b_k) \times \frac{g_k - b_k}{g_k + b_k} + t \cdot c_k$$

ここで、$d(g_k + b_k)$ 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

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

とする。ここで、$\tau$ はどの程度評価値を抑えるかを決める定数である。また、$s$ は閲覧者評価の割合、$t$ は機械評価の割合であり、$s + t = 1$ とし、機械評価の精度にあわせて t の値を大きくする。また $e_k$ は、$-1 < e_k < 1$ の値をとる。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アノテーションの評価の平均を信頼度 $p$ として、

$$p = d(n) \times \frac{1}{n} \sum_{k=1}^{n} e_k$$

とする。これを元にして、その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。

$$r_k = \frac{\alpha \cdot p + \beta(g_k - b_k)}{\alpha + \beta(g_k + bk)}$$

これにより、信頼度 $r_k$ を求める事ができる。信頼度 $r_k$ 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。ここで、$\alpha$ はアノテータ評価係数、$\beta$ は閲覧者評価係数であり、$\alpha + \beta = 1$ である。どれくらいアノテータに権威を持たせるかに応じて $\alpha$ の値を調整する。

アノテータに対する信頼度を計算する理由は、機械的にアノテーションを評価するのは難しい事、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事、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、信頼される情報を書き込むように暗黙的に強制させるところにある。

## 5. アノテーションを用いた応用

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・簡約に関する知的なシステムを作成した。自動解析から得られる結果のみに基づいて行うにはいずれにも難しい問題であるが、本研究で得られたアノテーションを用いることにより比較的短期間でシステムの構築が可能であった。

### 5.1 自然言語による Web 検索

カットとオブジェクトに対するカラーヒストグラム情報やテキストアノテーション情報、及びカットとオブジェクトの存在時間範囲を元にして、コンテンツの内容に基づく検索システムを試作した。本研究の特徴としては、人間が介在するアノテーションと、基本的に全自動解析されるヒストグラム情報やカット検出結果を用いて検索を行っており、人間と機械の両方のアノテーション結果をうまく使い分けている点がある。

まず、検索キーワードから茶筌 [5] を用いて動詞・形容詞・名詞・未知語を取り出す。未知語とは茶筌に登録されていない単語であり、名詞や英語などの場合が多い。次に形容詞や名詞から色にあたる単語（たとえば、赤い・黒い・青い・暗い・明るい等）を色名詞として、各シーンの色ヒストグラムを利用して検索結果を絞りこむ。また、色名詞以外の名詞・形容詞・動詞は、アノテーション情報に記述されたテキスト情報の記述との部分もしくは完全一致に応じて加点する。なお、アノテーションの種類が、名称 > 状況説明 > 補足説明 > 感想の順に重みをつけている。検索文に「おもしろい」、「悲しい」などの印象アノテーションに関連する言葉が入っている場合は、それに対応する各シーンの印象アノテーションの値に応じて加点する。最後に、全ての印象アノテーションの合計値を各カットに割り当て、なんらかの興味のあるシーンとして、それに応じて加点する。

このようにして加点された得点に応じてソートし、順位づけをした上でユーザに Web ブラウザを使って検索結果 (図 5) を示した。

### 5.2 アノテーションによるビデオ簡約

印象アノテーションとテキストアノテーションの情報を用いて任意の長さにビデオコンテンツを簡約するシステムを作成した。「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簡約化を行っている。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

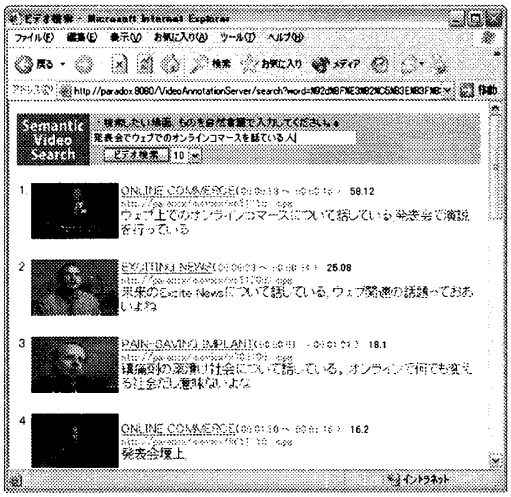


図 5: Web ビデオ検索結果

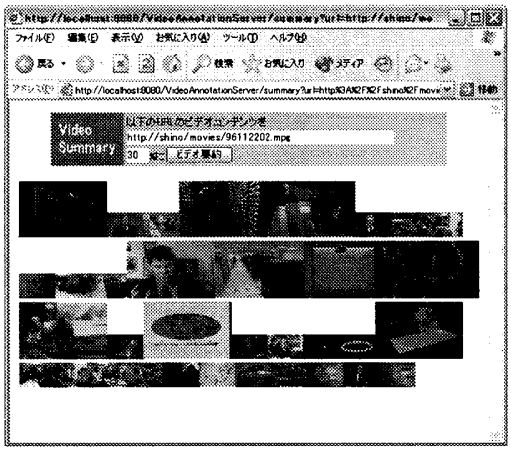


図 6: ビデオ簡約の例（小さいサムネイル画像の部分のカットが省略される）

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、そのカットの重要度と判定し、指定した時間を超えるまで、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、簡約とする。簡約結果を図 6 に示す。

## 6. 考察

本論文によってオンラインビデオコンテンツへの閲覧者によるアノテーションが可能であり、それによって各種応用が可能な事を示した。しかしながら、アノテーションに関する客観的な評価は基準が不明確で難しい。そこで、大学生 13 人に使ってもらい、Web 上での評価実験を行った。評価項目は以下の 5 点である。

- テキストアノテーションの使いやすさ
- 印象アノテーションの使いやすさ
- 思ったとおりに正確にアノテーションできるか
- 取っ付きやすさ

表 2: 評価実験 (被験者 13 人による 5 段階評価)

| 評価項目 | 5 | 4 | 3 | 2 | 1 | 平均 | 分散 |
|---|---|---|---|---|---|---|---|
| テキストアノテーション | 0 | 7 | 4 | 2 | 0 | 3.38 | 0.59 |
| 印象アノテーション | 2 | 8 | 0 | 3 | 0 | 3.69 | 1.06 |
| 正確にできたか | 0 | 4 | 9 | 0 | 0 | 3.31 | 0.23 |
| 取っ付き易さ | 4 | 8 | 0 | 1 | 0 | 4.15 | 0.64 |
| iVAS を利用したいか | 3 | 5 | 4 | 1 | 0 | 3.77 | 0.85 |

- iVAS を使ってアノテーションをしたいか

実験結果を表 2 に示す。テキストアノテーションはおおむね良好であるが、普通に掲示板感覚で書き込むよりは手間がかかるのが気になるという意見が多かった。また印象アノテーションは、使いやすいという意見が多い一方、どのタイミングでボタンを押していいのか戸惑うという意見もでた。また、正確にアノテーションできるかという点では、選択したい項目が存在しなかった時に困るという意見が多く、改善すべき点である。取っ付き易さ、アノテーションしたいかという質問に関しては、比較的良好な結果を得ており、本システムの有用性が確認できたと考えられる。また、現在では自分の記述した内容の編集ができず不満に思っている人も多い。

## 7. まとめ

本研究では、オンラインビデオコンテンツに対して、Web ブラウザを用いて閲覧者による不特定多数のユーザでのアノテーションを行う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、実際に得られたアノテーション情報を用いて各種の応用例を示した。

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事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、極めて発展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

[1] M. Davis. An iconic visual language for video annotation. In *Proceedings of IEEE Symposium on Visual Language*, pp. 196–202, 1993.

[2] Ching-Yung Lin, Belle L. Tseng, and John R. Smith. Videoannex annotation tool. http://www.research.ibm.com/VideoAnnEx/, 2002.

[3] Katashi Nagao, Shigeki Ohira, and Mitsuhiro Yoneoka. Annotation-based multimedia summarization and translation. In *Proceedings of the Nineteenth International Conference on Computational Linguistics(COLING-2002)*, Vol. 19, 2002.

[4] 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, 2002.

[5] 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.