

# Webブラウザを用いた閲覧者による実時間ビデオアノテーション\*

山本 大介

名古屋大学 情報科学研究科 メディア科学専攻  
yamamoto@nagao.nuie.nagoya-u.ac.jp

長尾 確

名古屋大学 情報メディア教育センター  
nagao@nuie.nagoya-u.ac.jp

## 1 はじめに

動画コンテンツの意味的な検索や要約のためには、そのコンテンツへのメタ情報（アノテーション）の付与が不可欠である。しかしながら、一般に動画に対する詳細なアノテーションは非常にコストが高く、作成に時間がかかる。そこで、いかにアノテーションコストを下げるかという観点から、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やり方で動画に対して実時間でのアノテーションを行うシステムが有用である。また、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）が実現できると思われる。

## 2 Web ブラウザによる実時間ビデオアノテーション

インターネット上から閲覧可能であるほとんど全てのビデオコンテンツに対し、一般的な Web ブラウザを用いてアノテーションの付与とその利用を行うシステムとして、intelligent Video Annotation Server (iVAS) を構築した (図 1)。ビデオコンテンツはインターネット上の任意の場所にあり、Registration Server を用いて登録する。登録されたコンテンツは Video Analysis Server でカット検出を行い、サムネイル画像と、時間情報をデータベースに保存する。これらの情報を元にして閲覧者は Web ブラウザを用いてアノテーションを行い、Video Annotation XML Database に格納される。

ブラウザの画面は図 2 のようになる。印象アノテーション (左)、動画 (中上部)、テキストアノテーション一覧 (中下部)、サムネイル画像を用いたスクロール可能なシークバー (右) になる。

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

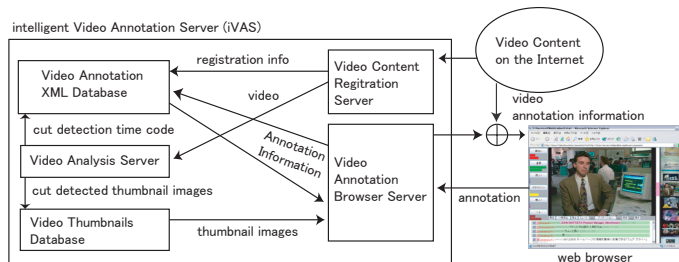


図 1: システム構成図



図 2: ブラウザ画面

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ー上に保存するプログラムとしてカット検出サーバーも作成した。カット検出のアルゴリズムには、分割 二乗検定法 [1] を採用した。

テキストアノテーションは、任意の連続するシーンに対して、テキストでアノテーションする方式である。全てのアノテーションには、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・人・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) も選択できるようにした (図 3)。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、閲覧者によってアノテーションの評価が行えるようにした。

また、シーンごとに他の閲覧者がアノテーションした

\*Web-based Realtime Video Annotation by Audience by Daisuke Yamamoto (Graduate School of Information Science, Nagoya University) and Katashi Nagao (Center for Information Media Studies, Nagoya University)

情報をリアルタイムに表示させるインタフェースを作成し、シーンに対する補助的な情報を得ることができるように工夫している(図2の下部)



図3: テキストアノテーション

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感情(例えば、面白い・つまらない・重要である)などをボタンを押すだけでアノテーションできる仕組みである。より面白いシーンでは「面白い」ボタンをすばやく連打することによって行う。ボタンを押すことによりその瞬間の前後に対し、正規分布を  $N(\mu, \sigma^2)$  とすると以下の式でパラメータを付与する。

$$\sum_{i=\text{全ての印象アノテーション}} N(t_i, m)$$

ただし  $t_i$  は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。

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、それぞれ任意の印象ボタンを作成できる。

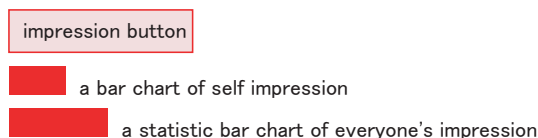


図4: 印象アノテーション

### 3 アノテーションに対する信頼度

不特定多数のユーザーによるアノテーションを扱うためには、アノテーションに対する信頼度を考慮する必要がある。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」という原則を元にして以下の方法で計算している。

あるアノテーション  $A_k$  に対する信頼度は以下のように求める。まず、 $A_k$  に対する評価  $e_k$  をそのアノテーションに (good) の評定をした人の数  $g_k$  と × (bad) の人の数  $b_k$  により以下のように求める。

$$e_k = d(g_k + b_k) \times \frac{g_k - b_k}{g_k + b_k}$$

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、

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

とする。 $\tau$  は常数。

次に、アノテーションを行う個人(ここではアノテータと言う)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  $r_k$  を求めた。 $-1 < r_k < 1$  であり、数字が大きいほど、信頼度が高い。

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

とする。ただし、 $\alpha$  はアノテータ影響係数、 $\beta$  は評価の影響係数とする。これにより、信頼度  $r_k$  を求める。

## 4 おわりに

動画像へのアノテーションに関する研究には様々なものがある[2, 3]が、一般にアノテーションにかかるコストが高い。また、Webブラウザを用いてビデオコンテンツに注釈をつけようという試みはEmond[4]らのものがあるが、高度なアノテーションとしての利用を考えるとはいない。そこで、本研究では、Webブラウザを用いて「多数の閲覧者から少しづつ情報を集め、全体として有用な知識を得る」ことにより、少ないコストで動画像にアノテーションができないかと考え、そのために問題となる、閲覧者にとっての負担、また付与された情報の正確さの問題に対し、直感的なインタフェースの構築と、信頼度の計算を行うことにより一定の解決策を示したと考えている。

## 参考文献

- [1] 長坂晃朗, 田中謙. 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. 情報処理学会論文誌, Vol.33, No.4. 1992.
- [2] 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. 2003.
- [3] Katashi Nagao, Shigeki Ohira, Mitsuhiro Yoneoka. Annotation-based multimedia summarization and translation. In Proceedings of the Nineteenth International Conference on Computational Linguistics (COLING-2002). 2002.
- [4] Bruno Emond, Martin Brooks, Arnold Smith. A Broad-band Web-based Application for Video Sharing and Annotation. The Ninth ACM Multimedia Conference. 2001.