SceneCabinet：映像解析技術を統合した映像インデクシングシステム

谷口　行信[†]　南　憲一[†]　佐藤　隆[†]　桑野　秀豪[†]
児島　治彦[†]　外村　佳伸[††]

SceneCabinet: A Video Indexing System Integrating Video Analysis Techniques

Yukinobu TANIGUCHI[†], Kenichi MINAMI[†], Takashi SATOU[†], Hidetaka KUWANO[†], Haruhiko KOJIMA[†], and Yoshinobu TONOMURA[††]

あらまし　映像の効率的なブラウジングのためには，映像にインデックスを付与する作業が必要である．本論文は 5 種類の映像解析技術—ショット切換検出，カメラワーク検出，テロップ検出，音楽検出，音声（人の声）検出 —を統合した映像インデクシングシステム SceneCabinet について述べる．映像解析技術に基づく自動インデクシングアプローチは従来から提案されているが，自動付与できるインデックスは限られる上，その精度は 100%でない．このような映像解析の不完全性を補うために，本システムは自動付与されたインデックスを効率的に修正したり，関連情報を人手で簡単に付与するためのユーザインタフェースを提供する．具体的には，タイムラインと代表画像一覧を組み合わせることで，効率的にインデックスの修正や付与ができるようにする．本システムを用いると，手作業の場合に比べて約半分の作業時間でインデックス付与できることを評価実験により示す．更に，映像ストリーミング技術を統合した Web ページ作成ツールについて述べ，その応用事例を挙げることで本システムの有効性を示す．

キーワード　映像処理，インデクシング，ブラウジング，インタフェース

1. まえがき

映像アーカイブを構築する動きが近年盛んである．映像アーカイブの効率的な検索，ブラウジングには映像のインデクシング作業が欠かせない [4]．我々はインデクシング作業を効率化する目的でショット切換検出，テロップ検出などの映像解析技術を開発し，それらを核に映像インデクシングシステム SceneCabinet を開発してきた．

映像解析技術は盛んに研究されてはいるものの，自動付与可能なインデックスは限られる上，その精度は 100%ではない．したがって，自動付与されたインデックスを修正したり，関連情報を人手で付与するためのユーザインタフェース（以下，インデックス編集インタフェース）が実用システムには欠かせない．このような手作業のコストを低減することは大規模な映像アーカイブの構築に向けた重要課題である．

映像解析と手作業の情報付与を組み合わせる，半自動インデクシングのアプローチ自体は新しいものではない．しかし，映像解析によってどの程度，コストが低減されるかは明らかではなかった．

本論文は SceneCabinet システムの中で以下の 2 点に焦点をあてる．第 1 はインデクシング作業を効率化するためのインデックス編集インタフェースについてである．実験を通して作業効率の観点から本システムを定量的に評価する．第 2 は映像ストリーミング技術の統合についてであり，実際に博物館などで利用された事例を挙げることで本システムの有効性を示す．

本論文の構成は以下のとおりである．まず **2.** でシステム概要とインデクシング作業の流れを説明し，**3.** は実装した映像解析技術について，**4.** はインデックスの内部データ表現について述べる．**5.** は本論文の主題であるインデクシング編集インタフェースについて，**6.** はその効率評価について述べる．**7.** は映像ストリーミング技術の統合とその応用について述べる．

[†] NTT サイバーソリューション研究所，横須賀市
NTT Cyber Solutions Laboratories, 1–1 Hikarinooka, Yokosuka-shi, 239–0847 Japan
[††] NTT サイバーソリューション研究所，京都市
NTT Cyber Solutions Laboratories, Kyoto-shi, 619–0237 Japan

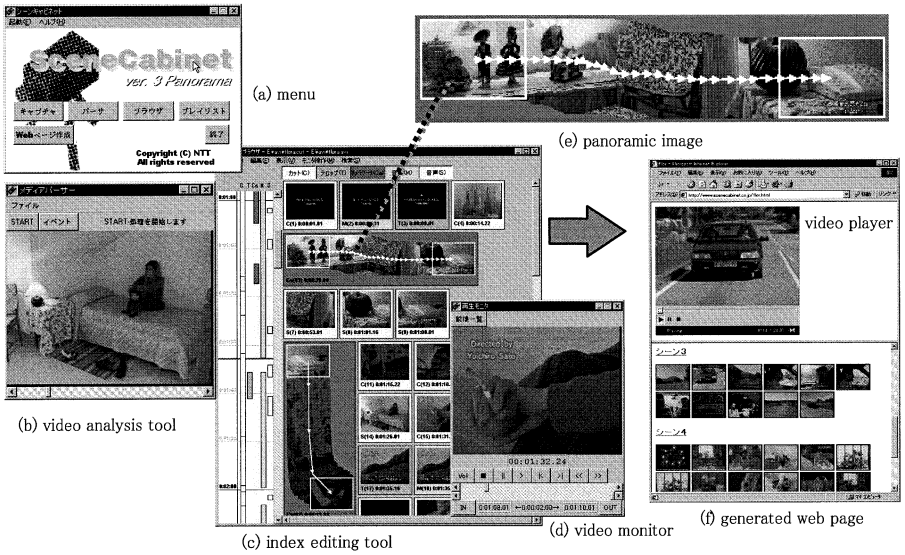


図 1　SceneCabinet システム概要
Fig. 1　SceneCabinet system overview.

2.　システム概要

2.1　機　　能

SceneCabinet を構成するツール群とその画面例を図 1 に示す．図 1 (a) に示すメニューから各種ツールを起動する．映像解析ツール（図 1 (b)）は映像から 5 種類のインデックス—ショット，カメラワーク，テロップ，音楽，音声—を自動検出する．詳細については **3.** で述べる．インデックス編集ツール（図 1 (c)）は解析結果をパノラマ画像（図 1 (e)）を含む代表画像として表示し，ユーザが簡単に誤検出を削除したり，コメントを付与できるようにしている．アイコンをマウスでクリックするとモニタ（図 1 (d)）に映像が再生される．インデックス情報は三つの形式で出力可能である．第 1 はデータベース管理システムへのデータ出力でありインデックス情報をリレーショナルデータベースの表に登録する．第 2 は印刷であり，映像の記録や打ち合せ資料としての利用を可能にする．第 3 は HTML ファイルとしての出力であり，**7.** で述べるようにストリーミング映像とリンクされた Web ページ（図 1 (f)）を制作することができる．

2.2　開 発 環 境

SceneCabinet は Microsoft Windows 上で動作するソフトウェアであり，使用プログラミング言語は C 言語と Tcl [7] である．スクリプト言語である Tcl を用いた理由は柔軟なカスタマイズとラピッドプロトタイピングを可能にするためである．処理速度が要求される部分（映像解析，映像入出力など）は C 言語を，出力部のように柔軟性が重視される部分は Tcl を用いて実装した．

3.　映像解析ツール

映像解析ツールは，入力として映像ファイルまたはビデオキャプチャボードからのデータを受け，5 種類のインデックス—ショット切換，カメラワーク，テロップ，音楽区間，音声区間— を検出する．処理可能な映像ファイル形式は MPEG-1, 2, AVI である．

ショット切換とカメラワークの検出には文献 [14] の方法を用いた．ショット切換検出の精度は 90%以上であり，ディゾルブやフェードも検出可能である．カメラワーク検出の精度はパン，チルトについて再現率で 90%以上である．ただし，ズームについては再現率がパンやチルトに比べて低い．テロップが表示されたフレームを検出するためにエッジペア特徴を用いた手法 [2] を用いており，ニュース映像 8 本を含む約 10 時間分の映像について実験を行ったところ約 95%のテロップを検出できた．音楽の検出は FFT スペクトラムが時間的に安定したピークをもつことを利用し，音声の検出は基本周波数の整数倍の高調波成分（ハーモニック構造）が現れる性質を利用して検出する [6]．こ

の方法の特長は音楽と音声が重なっていてもそれらを独立に検出できることであり，検出率は音楽と音声についてそれぞれ 90%, 80%以上である．

4. インデックスの内部データ表現

インデックスといってもシステムによって付与されるデータはまちまちである．本章は SceneCabinet が扱うインデックスを定義する．

映像に関するメタデータを表現するデータモデルとして，構造化モデル（structured model）[15] と層状化モデル（stratification model）[13] が代表的である．構造化モデルは映像をショット，シーン，ストーリの階層構造で管理し，記述単位としての区間は互いに重ならないように設定される．層状化モデルは重なりや包含を許す区間を記述単位とするため，より柔軟な記述が可能である．

本システムは層状化モデルを採用した．これはテロップ検出や音楽・音声検出の結果，得られる区間がショット区間を包含したり重なることがあるからである．インデックスを以下のように定義する：

$$I_i^j = (s_i, e_i; \mathrm{keyframe}_i, \mathrm{comment}_i),\ \ i = 1, 2, ...$$

ただし，s_i, e_i は区間の開始，終了時刻，$\mathrm{keyframe}_i$ は代表画像，$\mathrm{comment}_i$ はコメント（テキスト情報）である．また j はインデックス種別を表す変数でありショット，テロップ，カメラワーク，音楽，音声のいずれかである．映像解析では 5 種類のインデックスが自動付与されるが，初期値として $\mathrm{keyframe}_i$ には区間の先頭画像を，$\mathrm{comment}_i$ には空文字列を設定する．

5. インデックス編集ツール——インデクシング作業の効率化

本章では映像インデクシング作業の特性と問題点を明らかにし，インデックス編集ツールの実装について述べる．

5.1 映像インデクシング作業の特性

映像に対するインデクシング作業は静止画に対するそれと大きく異なる．

映像のインデクシング作業は次の三つのプロセスに分けて考えることができる：

α　内容把握：映像の画像・音声の内容を把握する，

β　区間指定：記述単位となる時間区間の開始，終了時刻を指定する，

γ　情報付与：キーワード，代表画像などを入力する．

サッカー映像を例にとって，ゴールシーンの情報をインデックスとして付与する作業を考える．映像を再生しながらゴールシーンを探すプロセスが α にあたり，ゴールシーンの始まりと終わりを指定するプロセスが β にあたり，ゴールを決めたチーム名を入力するプロセスが γ にあたる．

静止画の場合，画像メディアの多義性に関する問題はあるがプロセス α, β に要する時間は無視できる．一方，映像では α, β がコストの高いプロセスである；後述する評価実験で α, β が総作業時間の半分以上を占めることがわかった．これは，映像が時間軸をもったメディアであり，時間的に再生しないと内容を完全には理解できないという特徴をもつためである．したがって，プロセス γ だけでなく α と β を効率化することが重要な課題である．

5.2 従来技術

（1） 内容把握プロセス α の効率化技術．映像ブラウザが各種提案されているが，それらは映像の要点を見せることで内容把握プロセス α を効率化するものである．要点として，例えば，ショットの先頭画像をアイコンとして一覧表示する方法がある [15].

しかしインデクシング作業では映像を多面的に分析する必要があり，一面的な要点を示すだけでは不十分である．映像の概要に興味がある場合もあれば，詳細に関心がある場合もある．画像だけでなく，音声を聞く必要がある場合も多い．映像を多面的に把握するにはショットだけでなく複数のインデックスをユーザにわかりやすく提示することが有効である．

従来技術として，図 2 (a) に示すように奥行をもったアイコンを一覧表示するものがある（例えば [16]）．

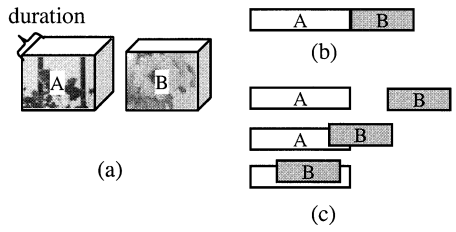
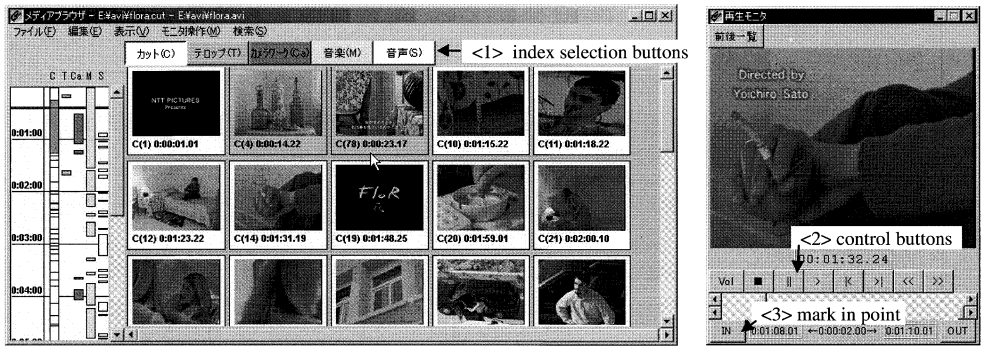


図 2　アイコン一覧表示と問題点：(a) アイコンの奥行として時間長を表現した例，(b) アイコンがショットを表している場合の区間配置，(c) 可能な区間配置

Fig. 2　Icon catalog representation and its problem: (a) the duration of a segment is represented with the icon depth, (a) the temporal arrangement of shots, (b) possible arrangements of segments.



(a) timeline (b) icon catalog (c) monitor

図 3　インデックス編集インタフェース
Fig. 3　Index editing interface.

これはアイコン一覧表示にタイムラインを組み合わせたもので，映像の概要とショット長を同時に表現している．アイコンがショットに対応する場合は図 2 (b) に示すように，アイコン A, B の時間的な配置が一意に定まる．しかし，アイコンがショット以外の音やテロップに対応する場合を考えると，図 2 (c) に示すような複数の配置があり得る．このように，複数のインデックスを同時に扱おうとすると，従来法では区間の時間的な関係を表現しきれないという問題がある．

一方，タイムライン表示（例えば [12]）だけでは映像の概要を把握しづらいという問題がある．

SceneCabinet はアイコン一覧表示とタイムライン表示を連携動作させることで，複数のインデックスを一画面に表示している．

（2） 区間指定プロセス β の効率化技術．映像をショットや，より意味的な単位であるシーンに分割する手法が開発されているが，いずれも 100%の精度ではない．

SceneCabinet では映像解析技術によりあらかじめ画像，音声の切れ目で映像を区間に分けておき，ユーザがそれらを併合したり再分割することで区間設定プロセス β を効率化する．

（3） 情報付与プロセス γ の効率化技術．近年，映像ロギングシステムとして商品化されているソフトウェアはプロセス γ を自動化するために，クローズドキャプション，文字認識技術，音声認識技術などを利用してテキスト情報を映像から抽出している [1]. CueVideo [8] はキーボード入力の代わりに音声認識技術を用いることで情報付与プロセスを効率化している．

SceneCabinet はキーワードやテキストの入力については特別な工夫をしていないが，後述するルール指定による代表画像自動選択機能により代表画像の効率的かつ柔軟な設定を可能にしている．

5.3　複数インデックスの同時表示

図 3 に SceneCabinet におけるインデックス編集インタフェースの実装を示す．本インタフェースは三つのビュー，(a) タイムライン部，(b) アイコン一覧部，(c) モニタ部から構成される．

タイムライン部（図 3 (a)）はインデックス区間の時間長や，時間的な関係（重なりや時間的な配置）を表現している．図 3 (a) のタイムラインは，左からショット，テロップ，カメラワーク，音楽，音声のインデックスに対応している．

アイコン一覧部（図 3 (b)）は代表画像（$\mathrm{keyframe}_i$）をアイコンとして一覧表示する．アイコンはインデックス種別により色分けされ，開始時刻 s_i の順に並べられる．一つのアイコンがタイムライン部の一つの方形に対応する．インデックス選択ボタンによりアイコンの表示・非表示を切り換えることで，ショット一覧，パノラマ画像一覧，テロップ画像一覧，音楽一覧，音声一覧を表示でき，多面的な映像分析を支援する．アイコンの上にマウスカーソルを移動させると対応する区間が動画表示される．

モニタ部（図 3 (c)）は映像の再生，停止，早送り，巻き戻しなどの操作を可能にする．

それぞれのビューは特に目新しいものではないが，本インタフェースの特徴は三つのビューが連携動作するようにした点にある．例えば，アイコン一覧部で一

表 1　ユーザインタフェース要素の役割分担
Table 1　Functions of the user interface elements.

	アイコン一覧	タイムライン	モニタ
a. 概要	○	×	×
b. 画像内容	○	×	○
c. 音声内容	×	×	○
d. 区間長	×	○	×
e. 区間の関係	×	○	×

つのアイコンを選択すると，タイムライン部の対応する方形がハイライト表示され，同時にモニタ部に対応する区間の先頭画像が表示されるようにしている．三つのビューには表 1 に示す役割分担があり，組み合わせて連携動作させることで多面的な映像把握を支援し内容把握プロセス α を効率化している．

講演映像を例にとって連携動作の効果を具体的に説明する．講演映像では講演者とスライドのショットを切り換えながら収録されることが多い．その場合，アイコン一覧（図 3 (b)）にはスライドと講演者のアイコンが並ぶ．このアイコン一覧を用いると，スライドのアイコンを手掛りに講演概要（表 1a）を把握でき，スライドのアイコンをクリックするとその画面がモニタ（図 3 (c)）に拡大表示されるのでスライドの内容も把握できる（表 1b）．タイムライン（図 3 (a)）を用いて，ハイライト表示された区間と重なりをもつ音声区間を見つけることができる（表 1 e）．音声区間に対応する矩形をダブルクリックすることで，話の切れ目から講義が再生される（表 1 c）．このようなインタラクションが可能になったのは三つのビューを連携動作させたためである．

5.4　アイコン操作によるインデックス編集

アイコンを追加，削除することで自動付与されたインデックスを修正していく．誤検出を削除するにはアイコンを削除すればよい(注1)．アイコンに対する操作として次の六つが定義されている：追加，削除，併合，開始終了時刻変更，代表画像変更，コメント入力．併合操作は以下のように定義される：$\mathrm{merge}(I_0, ..., I_n) = (\min(s_i), \max(e_i); \mathrm{keyframe}_0, \mathrm{comment}_0)$．アイコンの併合操作によりユーザが意味的な区間を効率的に指定できるようになる（区間指定プロセス β の効率化）．

5.5　ルール指定による代表画像自動選択

情報付与プロセス γ を効率化するために代表画像自動選択機能を提供する．

ユーザがメニューから代表画像選択ルールを選ぶと，代表画像がルールに従って置き換わる．ルールとして，ショットの先頭から x 秒後，末尾から x 秒前，ショットの中点，テロップ画像を代表画像とするものがある．ユーザはいろいろなルールを試行錯誤したり，組み合わせることで柔軟かつ効率的な代表画像選択を可能にする．なお，簡単なスクリプトを記述することでユーザがルールを定義できるようになっている．

6.　評 価 実 験

6.1　実験 1．タスク別作業時間の測定

提案システムを用いたインデクシング作業（半自動インデクシング作業 A）と手動インデクシング作業 M を作業効率の点から比較する．

6.1.1　実 験 条 件

実験には研究紹介用ビデオ 2 本（それぞれ約 4 分）とニュース（10 分）の計 3 本を用いた．フレームレートは 15 フレーム/秒，被験者はコンピュータの操作に習熟した男女，5 名である．映像解析には Pentium III 733 MHz のデュアルプロセッサ PC を用いた．

6.1.2　タスク設定

手動インデクシング作業 M は以下の二つのサブタスクからなる：

[M1] カット検出（手動）：映像に含まれるすべてのカット（漸次ショット切換を含む）を目視により検出し順次登録していく作業である．具体的な手順は以下のとおり：1) モニタ（図 3 (c)）を用いて映像を先頭から標準速で再生する，2) カットを見つけると再生停止ボタン（図 3<2>）を押す，3) コマ送り，コマ戻しのボタンを操作してカットの先頭画像を表示させる，4) IN ボタン（図 3<3>）を押下し Insert キーを押すとカット点が登録され，アイコン一覧部（図 3 (b)）にアイコンが一つ追加される，5) 1)〜4) の作業を映像の終わりまで繰り返す．ただし，すべての操作はマウスの代わりにキーボードでも行える．

[M2] 代表画像選択（手動）：タスク M1 の結果をショット一覧として表示しておき，ショット中にテロップが含まれるか調べ，存在すればその画像を代表画像として登録する．具体的な手順は以下のとおり：1) インデックス選択ボタン（図 3<1>）を操作してアイコン一覧（図 3 (b)）にショットに対応するアイコンのみを表示する，2) アイコンをダブルクリックしショットの先頭から映像再生する，3) テロップが表示された場

(注1)：ショットに対応するアイコンを削除した場合は直前のショットの終了時刻 e_i を調整することで，ショットとショットの間にタイムライン上ですき間ができないようにしている．

表 2　タスク別作業時間と成果物の品質
Table 2　Task times and the quality measures of products.

映像	長さ	ショット数	タスク別作業時間*					ショット再現率**	
			A1	A2	M1	M2	C1	半自動 A	手動 M
デモ 1	218 秒	26	.54	.44 (.12)	2.5	.81 (.16)	1.2 (.51)	100%	98%
デモ 2	240 秒	27	.54	.34 (.09)	2.1	.64 (.09)	1.7 (.47)	98%	100%
ニュース	600 秒	84	.68	.26 (.08)	2.6	.64 (.07)	.82 (.09)	94%	99%
総平均			.58	.35	2.4	.69	1.3		

*　5 名の被験者について作業時間（映像長に対する比）を平均した値，括弧内の数値は標準偏差を表す．
A1：映像解析，A2：修正（削除），M1：カット検出（手動），M2：代表画像選択（手動），C1：コメント入力．
** 再現率 R=（正しく検出できたショットの数）/（検出すべきショットの数）．

合はその時点で特定キーを押下し，その画像を代表画像として登録する．4) 2) と 3) をテロップが表示されていないすべてのアイコンについて繰り返す．

半自動インデクシング作業 A は以下の二つのサブタスクからなる：

[A1] 映像解析（自動）：映像解析ツールを用いて 5 種類のインデックスをすべて自動付与する．

[A2] 修正（削除）：タスク A1 の結果をインデックス編集ツールを用いてラフに編集することにより映像カタログを作成する作業である．ラフな編集とは誤検出されたショット，テロップを削除するものであり，アイコンの追加は行わないことにした．編集後，**5.5** に述べた代表画像自動選択機能を用いてテロップ画像を代表画像に設定する．アイコンを見ただけでは誤検出かどうか判断できない場合もあったので，映像を再生して確認してもよいと被験者に指示した．

更に，コメント入力タスクを以下のように設定した．

[C1] コメント入力：すべてのショットについて，代表画像に表示されたテロップのテキストをキーボードから入力する作業である．

上記タスクを設定するにあたり，手動で行うタスク（A1 以外）は個人差を減らすためにできるだけ主観が入らないように注意した．例えば，代表画像選択は実際は主観的なものだが，テロップの有無を基準に選択するように指示した．

6.1.3　実 験 結 果

上述した五つタスクの作業時間を 3 本の映像について測定した結果を表 2 に示し，平均作業時間を図 4 のグラフに示す．ただし，タスク M1 については 2 名の被験者についてだけ実験を行った．

コメント入力タスク C1 を含めた総作業時間は，手動インデクシングの場合は映像長の約 4.3 倍，半自動インデクシングの場合は約 2.2 倍であった．このことから映像解析技術を統合したことで作業時間が約 1/2

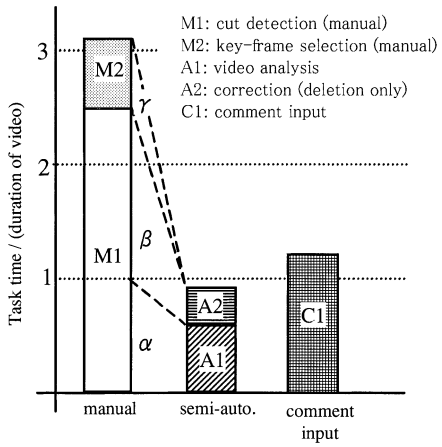


図 4　タスク別インデクシング作業時間
Fig. 4　Performance time for each indexing task.

に，コメント入力タスク C1 を含めない場合は約 1/3 に，短縮されることがわかった．映像解析 A1 は，例えば深夜にバッチ処理可能であるのでオペレータの稼動時間は更に短縮されることに注意する．

手作業でのカット検出 M1 はコストの高い作業であることがわかった．実時間の 2.4 倍の稼動がある．また作業後のインタビューで他のタスクより疲労の度合も高いことがわかった．タスク M1 を詳しく見ると，カットが見つかるまで映像を再生するフェーズ（内容把握プロセス α にあたる）と，再生を停止してからこま送りでカットの位置決めを行うフェーズ（区間指定プロセス β にあたる）からなる．ミスによる見直しがないと仮定すれば内容把握 α に要する時間は映像長に一致するはずで，残りが区間指定 β に費やされることになる．したがって，図 4 に示すように，プロセス α と β に費やされる時間の割合は 1 と 1.4 と考えることができる．一方，タスク A1，A2 がそれぞれプロセス α,β に対応するとみなす(注2)と，提案システ

ムを用いることで，手作業 M に比べてプロセス α,β がそれぞれ約 40%，70%だけ短縮された．カットの位置決めに要する時間を計算したところカット 1 個当り約 11 秒であった．

代表画像自動選択機能を利用しない場合は，タスク A1 と A2 に加えてタスク M2 を行う必要があるので，作業時間が映像長の約 0.7 倍だけ増加することが見込まれる．図 4 に示すように，タスク M2 は情報付与プロセス γ にあたる．

映像カタログの品質をショットの再現率により評価した．表 2 に示すように “デモ 1” を除いて半自動インデクシング作業 A は手動インデクシング作業 M と比べて再現率が低かった．原因は映像解析 A1 で検出もれがあったためである．

検出もれを減らし再現率を高めるには，1) 検出できなかったアイコンを手作業で追加する方法，と 2) しきい値を調整し映像解析 A1 の段階で検出もれを減らす方法がある．以下の実験 2, 3 では 2 通りの方法で作業時間がどのように変化するかを測定する．

6.2 実験 2. 追加タスクに要する作業時間

追加タスク A3 を以下のように設定した．

[A3] 追加：自動解析 A1 で検出できなかったアイコンを追加する作業である．具体的な操作は以下のとおり：1) アイコン一覧（図 3 (b)）にショットに対応するアイコンのみを表示する 2) アイコンをダブルクリックしショットの先頭から再生する，3) カットの検出もれを発見した時点で再生停止する，4) タスク M1 の 3), 4) の操作を行いアイコンを追加する，5) 2)〜4) をすべてのアイコンについて繰り返す．なお，被験者には操作 1) の途中で早送り再生に切り換えてもよいと指示した．

修正タスク A2 と追加タスク A3 をこの順に実行した場合の総作業時間，作業後のショット再現率，適合率を 5 名について測定した．適合率 P は $P=$（正しく検出できたショットの数）/（検出した数）で計算され誤検出の程度を表す．実験に使用した映像は，実験 1 で再現率が 94%と最も低かった “ニュース” である．

実験結果を図 5 (a) に示す．追加タスク A3 により，94%であった再現率を手動インデクシング作業と同程度にまで引き上げることができた．作業時間はすべてを手作業で行うよりは短いが実時間の 1.9 倍かかった．

6.3 実験 3. しきい値と作業時間の関係

ショット切換検出処理のしきい値を 3 段階（大，中，小）と変化させた．しきい値を下げていくと，検出処理 A1 のショット再現率は 94%, 98%, 100%と上昇するが，適合率は逆に 98%, 73%, 37%と下がって余計なアイコンが多く検出された．実験 2 と同様に “ニュース” を実験映像として使用し，修正タスク A2 に要する作業時間と修正後の再現率，適合率を測定した．ただし追加タスク A3 は行っていない．

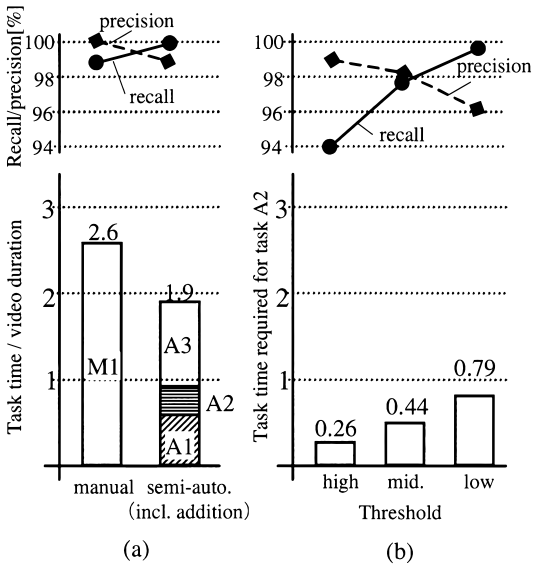


図 5 作業時間と品質の関係：(a) 手作業 M と半自動作業 A（追加あり），(b) ショット切換検出のしきい値を 3 段階に変えた場合

Fig. 5 Relation between task times and quality measures: (a) manual indexing M vs. semi-automatic indexing A (including addition), (b) at three thresholds used in the shot-change detection process.

実験結果を図 5 (b) に示す．しきい値を下げると作業時間が増加した．これは削除すべきアイコンの数が増え，作業量が増加するためである．適合率が下がったのは，削除すべきアイコンの数が増えたためミスが増加したものと考えられる．

6.4 考　　察

本評価実験をまとめると，提案システムは多少の検出もれが許容されるケースに特に有効であるといえる．映像ブラウジングへの応用（応用事例については次節

(注2)：プロセス α, β をそれぞれタスク A1, A2 に対応づけた理由は以下のとおりである：1) 映像解析 A1 は映像内容を人が確認する内容把握プロセス α を代行している；2) 修正タスク A2 は誤検出されたアイコンを削除することでショット区間を正確に指定しているので，区間指定プロセス β に相当する．ただし，厳密にいえば，修正タスク A2 でも映像の内容を確認することが頻繁にあり，タスク A2 とプロセス β を単純に対応づけることはできない．

で述べる）など多くの場合で有効であると考えている．実験 2 で示したように検出もれの追加が高コストである点は本インタフェースの問題点であり，その改良は今後の課題である．現状は，実験 3 に示したように検出処理のしきい値を低めに調整して検出もれを少なくすることで対処している．

草場ら [5] は手動インデクシングがコストの高い作業であること，カット検出処理の利用により作業時間を大幅に削減できることを実験により明らかにしている．本評価実験では新たに次のような知見を得た：1) 総作業時間だけでなくタスク別に作業時間を測定し，内容把握 α，区間指定 β に要するコストが高いことを示した；2) 実験 2, 3 を通して要求品質と作業時間がトレードオフの関係にあることを示した．

この評価実験では，ユーザインタフェース要素の一部しか評価できていない．実際には，プロトタイプ作成，主観評価，改良を繰り返しながらユーザビリティ向上を図った．**5.3** で述べたアイコンの動画表示も，プロトタイプを用いた主観評価により，検出もれの追加が難しいという上述した問題が明らかになったことから導入したものである．

7. Web ページ作成ツール——映像ストリーミング技術の統合

映像ストリーミング技術の進展に伴って，インターネットでの映像配信が身近になりつつある．蓄積型配信の場合，映像を先頭から再生できるだけではユーザの満足を得られなくなってきている．魅力的なサービスを提供するには，効率的なブラウジング手段，例えばニュースにはニュース項目を並べた，いわば目次ページ，を提供し，ニュース項目単位のきめ細かなアクセスを可能にすることが有効である．

多くの映像配信サーバは蓄積映像の途中再生を可能にする仕組みを備えているが，その機能を生かしきっていない．図 6 に示すように，途中再生のたびに Web ブラウザ，Web サーバ，映像配信サーバ，映像プレーヤの間で複雑なインタラクションがあり，HTML ファイル，メタファイル(注3)，映像データなど様々なデータが転送される．このように，目次ページを制作するには 1) 専門的な知識を要する，2) 手作業での区間指定にコストがかかる，という問題点があった．

SceneCabinet で管理されるインデックス情報は目次ページの制作に必要なすべての情報を含んでおり，その情報を埋め込むことで目次ページ制作支援ツールとして活用できる．

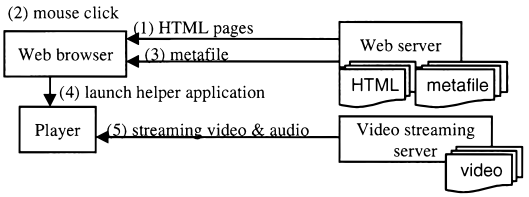


図 6　映像配信システムにおける途中再生の仕組み
Fig. 6　The mechanism of video streaming systems.

ただし，目次ページといっても用途によって求められるページデザインは様々であり，映像配信サーバによってメタファイルのフォーマットが異なるため，出力形式を固定することはできない．出力形式を簡単にカスタマイズできるように，柔軟性を考慮した設計が必要である．

7.1 実　　装

柔軟なカスタマイズのために，本システムはマクロを埋め込んだテンプレートファイルを複数用意しておき，マクロ展開により各種ファイルを生成する機構（テンプレート置換機構）を実装した．テンプレートにはファイル形式により，HTML 用，メタファイル用，JavaScript 用の 3 種類があり，メタファイルと JavaScript 用のテンプレートは映像配信サーバの種類ごとに用意した．映像配信サーバとして RealServer [9], SoftwareVision [3], Microsoft Windows Media に対応している．例えば，RealServer 対応のメタファイルテンプレートは

```
<%subst {$SC_URL?start=
  "[expr {$SC_START/1000.0}]"}%>
```

のようになり，インデックス情報に従って

```
rtsp://server/foo.rm?start=‘‘12.3’’
```

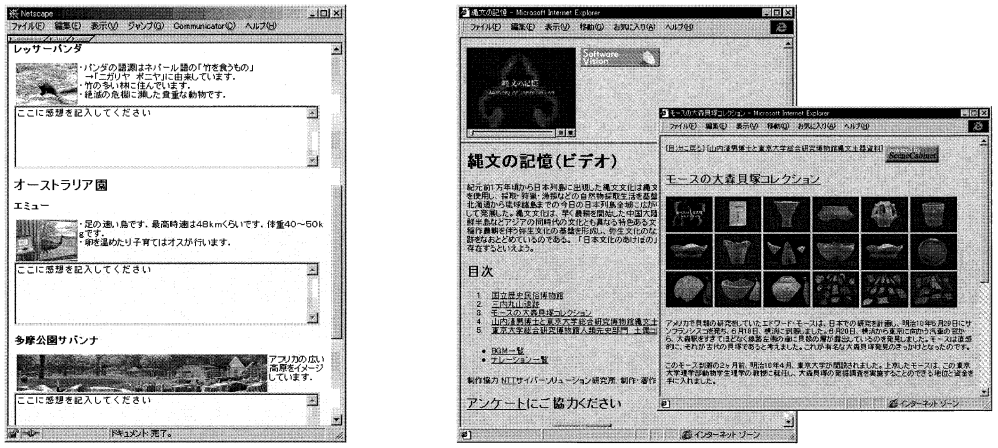
のように置換され ram ファイルが出力される．

7.2 応 用 事 例

映像ストリーミング技術の統合により可能となる応用例を示すことで，本システムの有効性を示す．

(a) 視聴ノート：視聴覚教育の補助ツールである視聴ノートの例を図 7 (a) に示す．本システムは Web サーバと生徒用の端末から構成され，先生が SceneCabinet を用いてコメント入りの視聴ノートを Web ページとしてあらかじめ作成しておき Web サーバに配置して

(注3)：メタファイルとは映像配信サーバの URL や再生開始点などを記述した小さなテキストファイルである（RealSystem では ram ファイルと呼ばれる [9]）．



(a) (b)

図 7 Web ページ例：(a) 視聴ノート，(b) デジタルミュージアム．
Fig. 7 Examples of the generated web pages: (a) video notes, (b) digital museum.

おく．生徒は映像を見た後で感想や疑問点を視聴ノートに記入して提出する．提出された視聴ノートは Web サーバで管理され，先生が感想が記入された視聴ノートを閲覧できる仕組みである．

(b) デジタルミュージアム：2000 年 3 月に東京大学博物館においてデジタルミュージアム 2000 [10] が開催された．その中で我々は SceneCabinet と映像ストリーミング技術 SoftwareVision [3] を組み合わせた実験を行い，SceneCabinet を目次ページ制作ツールとして使用した．制作した目次ページを図 7 (b) に示す．具体的には，60 分のビデオからパノラマ画像入りの映像カタログを自動作成し，テロップを手掛りに五つのパートに分けて構造化した．それぞれのパートは 18 から 144 個の代表画像を含んでおり，ナレーション原稿を付加情報として手作業で付加した．映像カタログを HTML ファイルに変換し Web サーバに，映像は SoftwareVision 形式 [3] に変換し映像配信サーバに配置した．映像を構造化したことで興味のある部分に素早くアクセスできるようになっている．このような目次ページを手作業で作成することも可能であるが，膨大な時間がかかるため非現実的であった．

映像ストリーミング技術により映像アーカイブがインターネットで公開できるという直接的な効果に加えて，映像を介したコミュニケーションを支援するツールとしても利用できることがわかった．視聴ノートは教育番組という素材を介した先生と生徒のコミュニケーションを支援する．また，社内ニュースをアーカイブした例では，社内での知識共有を支援するツールとして利用できることがわかった．

8. むすび

本論文では，映像インデクシングシステム SceneCabinet について以下の 2 点を中心に議論した．手作業によるインデクシングを効率化するためタイムラインとアイコン一覧を組み合わせたインデックス編集インタフェースを提案した．本論文の新規性は実装システムを用いてインデクシング作業に要する時間とその内訳を明らかにした点にある．映像ストリーミング技術を統合する際には，テンプレート置換機構により柔軟なカスタマイズを可能にした．更に，実際に博物館などの現場で応用された事例を示すことで本システムの有効性を示した．

マルチメディア情報の内容記述の枠組みを規定する MPEG-7 が標準化されつつあり，標準化により記述のインタオペラビリティが確保されコンテンツの流通利用が促進されることが期待される [11]．MPEG-7 の統合は今後の課題である．

謝辞 デジタルミュージアム 2000 において共同実験の機会を御提供頂いた東京大学の坂村健教授，越塚登助教授に感謝致します．有益な御助言を頂いた NTT サイバーソリューション研究所コンテンツ流通プロジェクト曽根原登プロジェクトマネージャ，NTT 東日本通信機器事業部浜田洋部長，阿久津明人氏，グループの皆様に感謝致します．

文　献

[1] C. Fuller, “Deploying Video on the Web,” in Web Techniques, Dec. 1999. （邦訳, “Web ビデオシステムの概要,” UNIX MAGAZINE, vol.15, no.3, 2000）

[2] H. Kuwano, Y. Taniguchi, H. Arai, M. Mori, S. Kurakake, and H. Kojima, “Telop-on-demand: Video structuring and retrieval based on text recognition,” Proc. IEEE International Conference on Multimedia and Expo 2000, pp.759–762, 2000.

[3] H. Jinzenji and K. Hagishima, “Real-time audio and video broadcasting of IEEE GLOBECOM ’96 over the internet using new software,” IEEE Commun. Mag., vol.35, no.4, pp.34–38, 1997.

[4] 児島治彦, “映像アーカイビング,” 情報処理, vol.41, no.6, pp.671–675, 2000.

[5] 草場匡宏, 高橋淳一, 洪　政国, “映像データベースにおける情報の入力と管理,” 情処学研報, 人文科学とコンピュータ, CH-15-2, pp.9–16, 1992.

[6] K. Minami, A. Akutsu, H. Hamada, and Y. Tonomura, “Video handling with music and speech detection,” IEEE Multimedia, vol.5, no.5, pp.17–25, 1998.

[7] J.K. Ousterhout, Tcl and the Tk Toolkit, Addison-Wesley, 1994.

[8] D. Ponceleon, S. Srinivasan, A. Amir, D. Petkovic, and D. Diklic, “Key to effective video retrieval: Effective cataloging and browsing,” Proc. ACM Multimedia ’98, pp.99–107, 1998.

[9] RealNetworks, RealSystem G2 Production Guide, 1999.

[10] 坂村　健（編）, デジタルミュージアム 2000, 東京大学博物館, 2000.

[11] 柴田正啓, “コンテンツ記述の標準化 MPEG-7,” 情報処理, vol.42, no.2, pp.176–182, 2000.

[12] 柴田巧一, 中屋雄一郎, 幸田恵理子, 山光　忠, 金田玄一, “業務用ネットワーク映像情報システム,” 日立評論, vol.81, pp.51–56, 1999.

[13] T.G.A. Smith, “Stratification: Toward a computer representation of the moving image,” A Working Paper from the Interactive Cinema Group, The Media Lab., MIT, 1991.

[14] 谷口行信, 阿久津明人, 外村佳伸, “PanoramaExcerpts：パノラマ画像の自動生成・レイアウトによる映像一覧,” 信学論（D-II）, vol.J82-D-II, no.3, pp.390–398, March 1999.

[15] Y. Tonomura, A. Akutsu, Y. Taniguchi, and G. Suzuki, “Structured video computing,” IEEE Multimedia, vol.1, no.3, pp.34–43, 1994.

[16] 上田博唯, 宮武孝文, 吉澤　聡, “認識技術を応用した対話型映像編集方式の提案,” 信学論（D-II）, vol.J75-D-II, no.2, pp.216–225, Feb. 1992.

（平成 12 年 8 月 25 日受付, 12 月 22 日再受付）

谷口　行信　（正員）

平 2 東大・工・計数卒．平 4 同大大学院工学系研究科修士課程了．同年日本電信電話（株）入社．現在 NTT サイバーソリューション研究所勤務．映像処理の研究に従事．平 12 本会論文賞受賞．情報処理学会，ACM 各会員．

南　憲一　（正員）

平 3 慶大・理工・電気卒．平 5 同大大学院修士課程了．同年日本電信電話（株）入社．現在 NTT サイバーソリューション研究所勤務．映像ハンドリングの研究に従事．

佐藤　　隆　（正員）

平 3 東大・工・電子卒．平 5 同大大学院情報工学・修士了．平 8 同博士了．同年日本電信電話（株）入社．現在サイバーソリューション研究所勤務．映像処理，映像データベース，映像インタフェースの研究に従事．工博．情報処理学会会員．

桑野　秀豪　（正員）

平 5 新潟大・工・情報卒．平 7 同大大学院修士課程了．同年日本電信電話（株）入社．現在 NTT サイバーソリューション研究所勤務．主に映像情報の構造化の研究に従事．

児島　治彦　（正員）

昭 57 早大・理工・数学卒．同年日本電信電話公社（現 NTT）入社．以来，手書き文字図形認識，文書画像理解，ISDN 端末システム，映像インタフェースの研究開発に従事．現在 NTT サイバーソリューション研究所主幹研究員．情報処理学会，IEEE 各会員．

外村　佳伸　（正員）

昭和 54 京大・工・電子卒．昭 56 同大大学院修士課程了．同年日本電信電話公社（現 NTT）入社．以来，画像を中心としたメディア変換装置の研究・開発，映像ハンドリングの研究に従事．昭 62〜63 米国 MIT メディア研究所客員研究員．現在 NTT サイバーソリューション研究所プロジェクトマネージャ．情報処理学会，映像情報メディア学会，IEEE, ACM 各会員．