

PanoramaExcerpts: パノラマ画像の自動生成・レイアウトによる映像一覧

谷口 行信[†] 阿久津明人^{††} 外村 佳伸^{††}

PanoramaExcerpts: Video Cataloging by Automatic Synthesis and Layout of Panoramic Images

Yukinobu TANIGUCHI[†], Akihito AKUTSU^{††}, and Yoshinobu TONOMURA^{††}

あらまし 映像の効率的な把握とアクセスを支援するためのブラウジングインタフェースは、映像ライブラリーなどの映像を扱うアプリケーションにとって重要な構成要素である。本論文は2種類のアイコン—パノラマアイコンと代表フレームアイコン—を一覧表示する映像ブラウジングインタフェース PanoramaExcerpts を提案する。パノラマアイコンはカメラ操作としてパン、チルト、又はズームを含む区間から画像列を合成することによって生成される。代表フレームアイコンはカット、ディゾルブ等のショット切換えを検出することによって各ショットごとに1枚ずつ選び出される。パノラマ画像を代表画像として用いることにより、代表フレームだけでは表現しきれなかったパン、チルト区間の全体像が表現できるようになる。更に、カメラの移動方向、速度などを可視化するアイコン表示方法についても提案する。PanoramaExcerpts を自動生成するために以下の処理を統合するアプローチを提案する: (i) ショット切換え検出方法, (ii) 安定なカメラ操作区間を抽出する方法, (iii) アイコンレイアウト方法。本論文では主に (ii), (iii) の処理について実験結果を含めて述べる。

キーワード パノラマ画像, 映像処理, Image Mosaic, ブラウジング, インタフェース

1. ま え が き

デジタルビデオ技術の普及に伴い、コンピュータで映像を扱うことが身近になってきている。映像情報の量が増加すればするほど、その中から本当に必要な情報を見つけ出すのが困難になる。映像の効率的な把握とアクセスを支援するブラウジングインタフェースは、デジタルビデオライブラリー、ノンリニア映像編集システムなどの映像を扱うアプリケーションにとって重要な構成要素である。

映像ブラウジングインタフェースの一つに、映像からその内容を代表する静止画像を抽出し一覧表示するものがある [2]。このインタフェースを設計する上で重要な点は、(i) 映像内容を効果的に表現した代表画像を抽出すること、(ii) 時間的な流れ (コンテキスト) を

利用者が直感的に把握できるように代表画像をレイアウトすること、(iii) 画面内にできるだけ多くの代表画像を表示し一覧性を高めること、である。

代表画像の抽出方法に関しては多くの研究がある。映像から場面の切り換わり (ショット切換えと呼ぶ) を検出し、ショットの先頭フレームを代表画像とする単純な方法 [1], [10], [15] のほかに、より効果的な代表フレームを選択するために動き解析 [11] や色特徴 [23] を用いる手法や、場面の繰り返しパターンを検出し冗長性を削減する方法 [5], [22] が提案されている。

従来アプローチの問題点の一つは、カメラ操作を含むショットからは代表フレームをうまく選択できないことである。図 1 (a) に示すショットはチルト (カメラを上下に動かす操作) を含むものであり、どの一枚を取っても代表フレームとして適切でない。

本研究では上記の問題点を解決するために、画像列を継ぎ目なく合成してできるパノラマ画像 (図 1 (b)) を代表画像として用いる。パノラマ画像はショットの全体像をコンパクトに表現している。従来のように代表フレームを“選択”するのではなく、代表画像を合

[†] NTT サイバースペース研究所, 横須賀市
NTT Cyber Space Laboratories, 1-1 Hikarinooka, Yokosuka-shi, 239-0847 Japan

^{††} NTT サイバースソリューション研究所, 横須賀市
NTT Cyber Solution Laboratories, 1-1 Hikarinooka, Yokosuka-shi, 239-0847 Japan

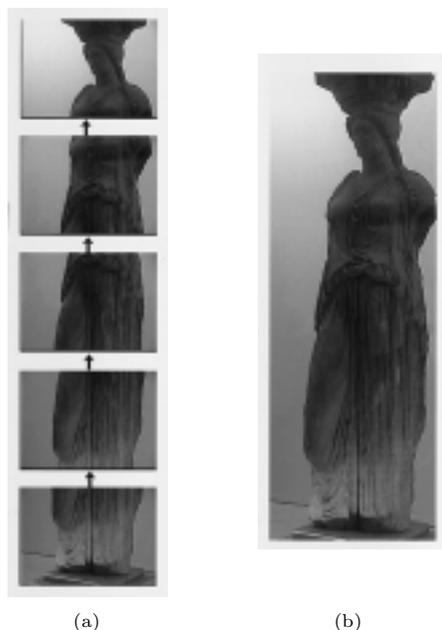


図1 代表フレーム選択アプローチの欠点: (a) 入力画像列, (b) パノラマ画像

Fig. 1 A limitation of keyframe extraction approach: (a) input image sequence, (b) the panoramic image created from (a).

成し“生成”するアプローチである。本論文はパノラマ画像を代表画像として用いた映像ブラウジングインタフェース PanoramaExcerpts と、それを自動生成するシステムを提案する。

パノラマ画像を代表画像として映像ブラウジングに用いるというアイデア自体は文献 [2], [7], [12], [17], [20] などに既に言及されている。Massey and Bender [7] は映像から、パノラマ画像を含むコミック本を作成するシステムについて述べているが、彼らのシステムはマニュアルでパノラマ画像を作成する場面を一つひとつ指定するものであり、我々の目的であるブラウジングインタフェースの自動生成には向かない。一般的な映像からパノラマ画像を自動抽出した例は我々の知る限りない。

パノラマ画像合成技術は Image Mosaic などと呼ばれ、バーチャルリアリティ [18], 画像圧縮 [4], 映像の2次利用促進 [7], 3D シーン再構築, 動物体認識 [14] など多くの応用をもち、多くの研究がある [3], [9], [17]。しかし、従来研究のほとんどが高解像度のパノラマ画像を生成することを目的としており“コントロールされた条件下で撮影された連続的な画像列”を入力とし

て仮定している。それに対して、本研究では“編集された長時間の映像”を対象とする。編集映像は被写体の動き、手ぶれなどのさまざまなノイズを含んでいるので、すべてのショットからカメラパラメータが正しく推定できるわけではないことが問題となる。また、推定の手がかりとなるエッジが映像中にない場合にも正しい推定はできない。したがって、代表画像にふさわしい見た目によりパノラマ画像を自動抽出するためには、手ぶれなどと区別して安定にカメラが操作された区間(安定カメラ操作区間と呼ぶ)を検出し、その区間に限定してパノラマ画像合成を行う必要がある。Smithら [13] は動きベクトルを利用して、パン、チルト、ズームを検出する方法を提案している。

パノラマ画像を導入することで新たに生じる問題点として、パノラマ画像の形状がカメラ操作によって変化するので、単純に並べて表示したのでは無駄な余白が生じて一覧性が低下することがある。本研究ではこの問題点を克服し一覧性を高めるために、時間順を保持した状態でアイコンをウィンドウ内に効率的にレイアウトするアルゴリズムについても提案する。

本論文の構成は以下のとおりである。2. は PanoramaExcerpts システムの概要について述べる。3. はパノラマアイコンを生成する処理について、4. はカメラの移動方向及び速度を可視化する方法について、5. はアイコンをウィンドウ内に効率良くレイアウトする方法について提案し実験結果を報告する。6. は PanoramaExcerpts システムについて、インタフェースと処理速度の観点から考察を加える。

2. PanoramaExcerpts システムの概要

PanoramaExcerpts インタフェースを自動生成する処理は大きく分けて二つのフェーズからなる。一つは、映像を解析しアイコンを抽出するフェーズ((1)~(5)), もう一つはアイコンをレイアウトしユーザに提示するフェーズ((6))である。処理の流れを図2に示す。

(1) ショット切換え検出: 入力映像からカット, ディゾルブなどのショット切換えを検出し, 映像をショットに分割する;

(2) カメラパラメータ推定: それぞれのショットに対してカメラパラメータを抽出する;

(3) 安定カメラ操作区間の検出: 安定カメラ操作区間を検出し, カメラ操作をパン, チルト, ズーム, その他の四つに分類する;

(4) パノラマアイコン合成: 安定カメラ操作区間

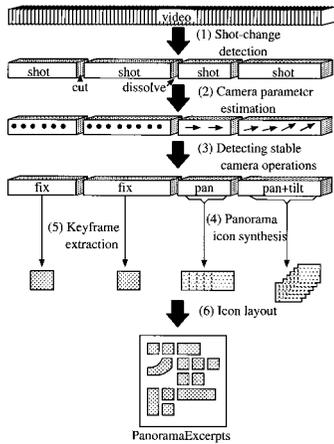


図2 PanoramaExcerptsシステムの概要
Fig. 2 Overview of the PanoramaExcerpts system.

のそれぞれについて、画像列を合成しパノラマイコンを生成する;

(5) 代表フレームアイコン抽出: 安定カメラ操作区間を含まないショットについては代表フレーム画像を各ショットから1枚ずつ抽出する. 現在のシステムでは各ショットの先頭画像を代表フレームとする;

(6) レイアウト: 無駄な余白が少なくなるようにウィンドウ内にアイコンをレイアウトし表示する.

本システムで特徴的なステップは(3)と(6)である. ステップ(3)は上述したように見た目のよいパノラマ画像を抽出するために必要である. ステップ(6)は一覧性を高めるために必要である. 3. ではステップ(2)~(4)について説明し, 5. ではステップ(6)について説明する.

なお, ステップ(1)のショット切換え検出法としては文献[16]の方法を用いた. ディゾルブを含めたショット切換えを検出できる方法である.

3. パノラマイコンの生成

パノラマイコンを生成する処理は次の四つのステップからなる: カメラパラメータ推定, 安定カメラ操作区間の検出, カメラ操作の分類, パノラマ画像合成. それぞれのステップについて以下の節で説明する.

3.1 カメラパラメータ推定

映像中の2枚の隣り合う画像を $f(x, y)$, $f'(x, y)$ とする. パン, チルト, ズームの三つのカメラ操作を説明するために, 次の座標変換

$$(x', y') = (ax, ay) + (d_x, d_y), \quad (1)$$

をカメラモデルとする. ここで, (x, y) は画像 f の点を表し, (x', y') は点 (x, y) の画像 f' 上における対応点を表す. a, d_x, d_y はそれぞれズーム, パン, チルトを説明するパラメータであり, まとめて $\theta = (a, d_x, d_y)$ とおく. 座標系の原点はフレームの中心とする. 式(1)は焦点距離が大きい場合にパン, チルト, ズームのカメラ操作を近似的にモデル化する. 本研究では精度よりも処理速度を重視しているため, 比較的単純なカメラモデルを用いている.

2枚の画像 f, f' の間の誤差尺度として, 最小2乗誤差 (Mean Square Error) を

$$MSE(\theta) = \frac{1}{N} \sum_{(x,y)} \{f(x,y) - f'(x',y')\}^2, \quad (2)$$

と定義する. ここで, 和は重なりをもつすべての画素についてとり, N は重なりをもつ画素の数である^(注1). $MSE(\theta)$ を最小化する θ をカメラパラメータとするわけであるが, 最適化には coarse-to-fine 法[4]を用いる. 具体的には, 2枚の画像 f, f' からそれぞれピラミッド画像を構成し, 低解像度の画像を用いて粗い推定を行い, 高解像度の画像を用いて精度を上げる.

3.2 安定カメラ操作区間の検出

時間区間 $I = [s, e]$ が以下の三つの条件をすべて満たす場合に区間 I を安定カメラ操作区間とみなす.

- 継続時間: カメラ操作は一定時間 D 以上継続すること. すなわち $e - s \geq D$;
- 滑らかさ: カメラ操作は滑らかであること. すなわち $\hat{\theta}_t \cdot \hat{\theta}_{t-1} > 0, t \in I$. ここで, $\hat{\theta}_t$ は時刻 t におけるカメラパラメータ推定値を表す;
- カメラモデル適合性: MSE の最小値 $MSE(\hat{\theta}_t)$ が $MSE(0)$ よりも十分小さい, すなわち

$$MSE(\hat{\theta}_t)/MSE(0) < T, \quad t \in I \quad (3)$$

ただし, T はしきい値であり $0 \leq T \leq 1$ である.

3.3 カメラ操作の分類

安定カメラ操作区間 I について, x 方向及び y 方向の総移動量, 総ズーム量を求め, それらの値があるしきい値以上の値をとる場合にパン, チルト, ズーム成分があると判定する.

(注1): 具体的には, 画像 f 内のすべての画素 (x, y) について, 式(1)により対応点 (x', y') を求め, その点が画像 f' の上に乗っていれば式(2)に従って和をとる. N は和をとった画素の数である.

表 1 カメラ操作分類の適合率 P 、再現率 R 、 T 、 M 、 F はそれぞれ正解数、検出もれ数、誤検出数を表す

Table 1 Precision P and recall R measures of the camera operation classification method. T , M , and F denote the numbers of camera operations, false negatives, and false positives, respectively.

source	pan			tilt			zoom		
	T	$M(R)$	$F(P)$	T	$M(R)$	$F(P)$	T	$M(R)$	$F(P)$
Greece	33	2(94%)	0(100%)	18	0(100%)	0(100%)	6	2(67%)	0(100%)
CNN	8	2(75%)	0(100%)	5	0(100%)	0(100%)	11	2(82%)	2(82%)
India	24	0(100%)	2(92%)	17	2(88%)	1(94%)	12	6(50%)	0(100%)

3.4 パノラマ画像合成

安定カメラ操作区間のそれぞれについてパノラマ画像を合成する。推定されたカメラパラメータを式 (1) に代入した変換式によって画像列を変換し、重ね合わせていくことで、継ぎ目のないパノラマイコンを合成する。重なった部分は上書きされる。ズームを含む場合、最も広角の（ズームアウトした）フレームを基準の大きさとして画像合成する。これは、合成されるパノラマイコンの大きさをほぼ同じにそろえるためである。

3.5 実験

3種類の映像（それぞれ 15 分程度）に対して、人手で抽出した安定カメラ操作区間と提案法で抽出したものを比較した結果を表 1 に示す。1 秒以上継続するカメラ操作区間のみを検出対象とし、パンとチルトが重複している場合にはパン、チルトの両方でカウントした。一つのカメラ操作区間を二つ以上の区間として認識した場合は誤検出として計数した。適合率 P と再現率 R を

$$P = (T - M - F) / (T - M), \quad R = (T - M) / T,$$

と定義する。ここで、 T 、 M 、 F はそれぞれ正解数、検出もれ数、誤検出数である。表 1 に示すようにパン、チルトについては適合率 90% 以上で検出ができた。検出もれは、例えば、画面内で大きな面積を占める被写体をパンで追っている場面で生じた。ズームに関しては検出もれが多く再現率が低くなった。これはズームによる画面変化が緩やかで式 (3) の条件を満たさなかったためである。

カメラモデル適合性の条件（式 (3)）は誤検出を抑える効果があった。この条件を除いて映像データ “Greece” に対して同様の実験を行ったところ誤検出が 8 個増加した。

不適切なパノラマイコンが抽出される原因としては以下のようなものがあった。ショット切換えの検出

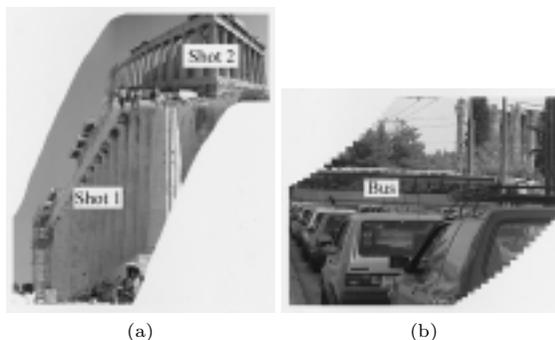


図 3 不適切なパノラマイコンの例：(a) 二つのカメラ操作区間を一つとみなし、二つのショットを合成した例、(b) フレーム境界で被写体（バス）の動きが存在したためバスが実際より長く表示された例

Fig. 3 Examples of inappropriate panorama icons: (a) a panorama icon wrongly generated by combining two consecutive shots, (b) a panorama icon in which a bus is longer than the actual one, due to the motion.

もれが原因で、ショット切換えをはさんで存在する二つのカメラ操作区間を一つとみなし、二つのショットを無理に合成して一つのパノラマイコンとして提示した例があった。図 3 (a) はディゾルブの検出もれが原因で、同じ方向にパンしている二つのショットを合成してできたパノラマイコンである。逆に、一つのカメラ操作区間から二つのパノラマイコンを抽出した例があった。これはカメラ操作の途中でカメラ操作が中断されたと判定されたことが原因であった。このように、本システムでは安定カメラ操作区間の誤検出が検出もれよりも悪影響を及ぼすので、適合率が高くなるようにしきい値 T を調整することが望ましい。

被写体の動きが存在する場合には、次のような問題が発生する可能性がある。(i) カメラパラメータ推定の誤差が大きくなること、(ii) 合成されたパノラマ画像の継ぎ目が目立つこと。(i) は上述した適合性条件（式 (3)）によって被写体の動きが激しい区間は安定カメラ操作区間でないと判定されたので、パノラマイ

コンは生成されなかった。(ii)に関しては、フレームの境界付近で被写体が動いている場合に継ぎ目が目立つ場合があった。図 3 (b) に示すパノラマアイコンでは、パンの最中にバスがフレームを横切ったため、バスが実際の長さより長く表示された。

4. カメラ操作軌跡の可視化

前章の方法で生成されるパノラマアイコンは、その形状によってカメラ操作のタイプを表現しているが、カメラの移動方向及び速度に関しては表現されていない。映像編集用素材データベースの検索などの応用では、各映像クリップの中でカメラがどのように操作されているかがキーとなる。そこで、カメラパラメータ推定ステップで得られた情報をパノラマ画像上に表現することが有効であると考えられる。

PanoramaExcerpts システムは、カメラ操作の分類結果に応じてカメラ操作の可視化方法を切り換える。ズームを含まない映像区間に対しては図 4 (a) に示すように画像中心の軌跡を矢印で表示し、ズーム操作を含む場合には、図 4 (b) に示すように画像の四隅の軌跡を表示する。

カメラ操作の可視化に関する従来研究としては、アイコンの奥行方向にカメラ操作情報をもたせる VideoSpaceIcon [20] や、代表フレームの上に矢印によってカメラ操作を表現する方法 [21] が提案されている。

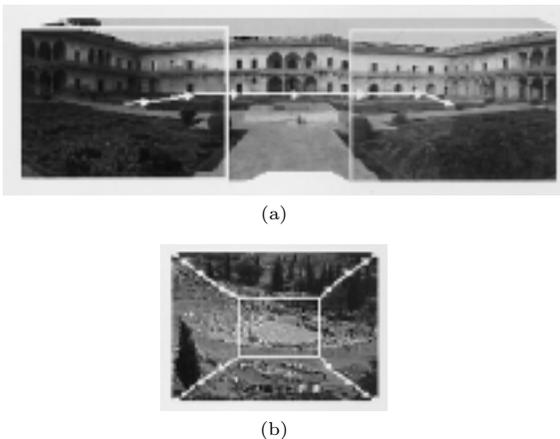


図 4 カメラ操作軌跡の可視化：(a) ズームを含まない場合，(b) ズームを含む場合

Fig. 4 Visualization of camera operations: (a) for shots without zoom operation, and (b) for shots with zoom operation.

これに対し提案方法は、パノラマアイコンによって撮影された空間の全体像をコンパクトに表現し、その上にカメラ操作の方向、速度を表現した矢印を重畳したことで、カメラ操作情報と映像内容を同時に直感的に表現している点が特長である。

5. アイコンレイアウト

アイコンレイアウトに関する従来研究として、ウィンドウ内に単純に一覧表示する方法 [19], [21]、階層的表示を行う VideoMagnifier [8]、アイコンの優先度を考慮してアイコンの大きさを変化させる VideoPoster [22] などがある。アイコンを取捨選択したり大きさを変化させることでブラウジング効率を高めようとする試みである。

ここではアイコンを変形しない基本的なレイアウト方法について検討する。本論文で導入したパノラマアイコンは従来と異なり、その大きさ、形状がカメラ操作に応じて変化するため、非定型のアイコンを効率良くレイアウトすることが一覧性を高めるために必要である。

単純なレイアウト方法(方法 A と呼ぶ)として、左から右へ順にアイコンを配置していき、右端に達したら改行するというものが考えられる。しかし図 5 (a) に示すように縦長のアイコンが存在すると無駄な余白が生じてしまう。

提案法は無駄な余白を減らすために、図 5 (b) に示すように 1 行を複数行に再分割することを許すようにする。

無駄な余白を最小化する、いわゆるパッキングアルゴリズムに関して多くの研究があるが [6]、それらのほとんどは VLSI 設計等の非対話的な応用のために設計されている。従来研究と異なる点は、(i) アイコンの時間順(コンテキスト)を保存する必要があり、(ii) 対

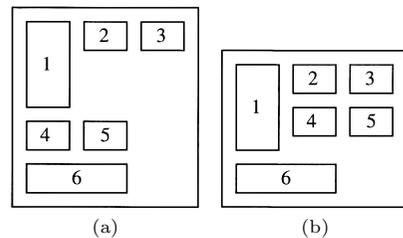


図 5 レイアウト：(a) 方法 A，(b) 提案法
Fig. 5 Layout generated by (a) a simple method A, and (b) our proposed method.

話的システムのため高速性が要求されることである。

5.1 レイアウト手続き

本節は提案法の具体的な手続きについて述べる。簡単のためアイコンに外接する方形 $R_i, i = 1, \dots, N$ を幅 W , 高さ不定のウィンドウ内にレイアウトする問題を考える。ただし, N はアイコン数を表し, R_i はウィンドウの縦横の辺に平行な方形とする。時間的順序が保存されるという制約の下で, 無駄な余白を最小化するように方形を左上から順に配置していく。

具体的な手続きは以下のとおり。

1. 他の方形によって占有されている領域(占有領域と呼ぶ)を初期化する: $O = \phi$. 占有領域 O は,

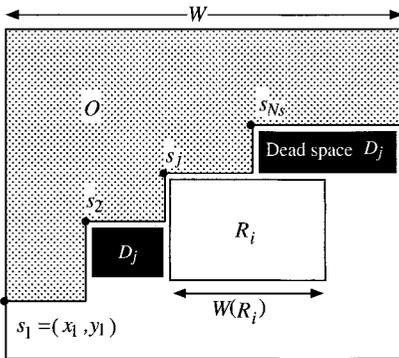


図 6 方形 R_i を j 番目の角 (x_j, y_j) に配置した場合に生じる無駄領域 D_i
 Fig. 6 Dead space D_i generated by placing rectangle R_i justified at j -th step corner (x_j, y_j) .

図 6 に示すように階段状の領域となるので, 階段の角の座標値 $s_j = (x_j, y_j), j = 1, \dots, N_s$ によって表現できる。ただし, ウィンドウ左上隅を座標原点とする。 $(x_1, y_1) = (0, 0), N_s = 1$ と初期化する。

2. 各方形 $R_i, i = 1, 2, \dots, N$ について:

2.1. 各階段の角 $(x_j, y_j), j = 1, 2, \dots, N_s$ について: 方形 R_i を j 番目の角 (x_j, y_j) に配置することによって生じる無駄な領域を D_j とする。図 6 に示すように, 無駄領域 D_j は二つの領域からなり, 方形を今後配置できなくなる領域である。もし方形 R_i を配置するスペースがない場合, すなわち $W(R_i) > W - x_j$ ならば便宜的に $D_j = \infty$ とおく。ただし, $W(R_i)$ は方形 R_i の幅を表す;

2.2. 無駄領域 D_j の面積を最小化する配置 j^* を求める: $j^* = \arg \min_j |D_j|$, ただし, $|D_j|$ は領域 D_j の面積を表す;

2.3. 方形 R_i を j^* 番目の角に配置する;

2.4. 占有領域を更新する: $O = O \cup R_i \cup D_{j^*}$. それに伴って占有領域の形状が変化するので (x_j, y_j) と N_s を更新する。ステップ 2.1 に戻る。

5.2 実験

単純なレイアウト方法と提案法とを比較するためにパッキング率 PR を $PR = H(\text{提案法})/H(\text{方法 A})$ と定義する。ここで, $H(\text{提案法}), H(\text{方法 A})$ はそれぞれ提案法と方法 A によって出力されるレイアウトの高さを表す。パッキング率が 1 より小さい場合は提案法

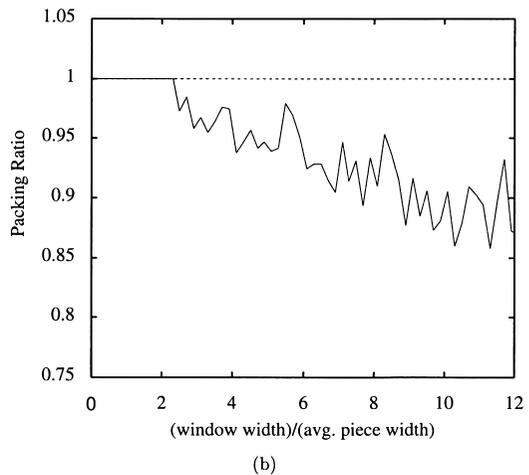
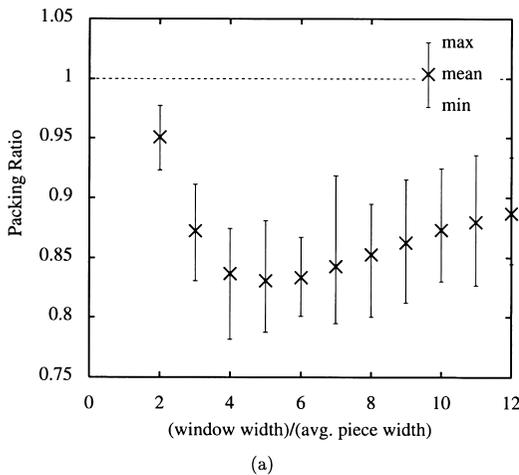


図 7 パッキング率: (a) 一様乱数により生成されたデータセット, (b) 実データセット
 Fig. 7 Packing ratio measured on: (a) randomly generated datasets, and (b) a real dataset.

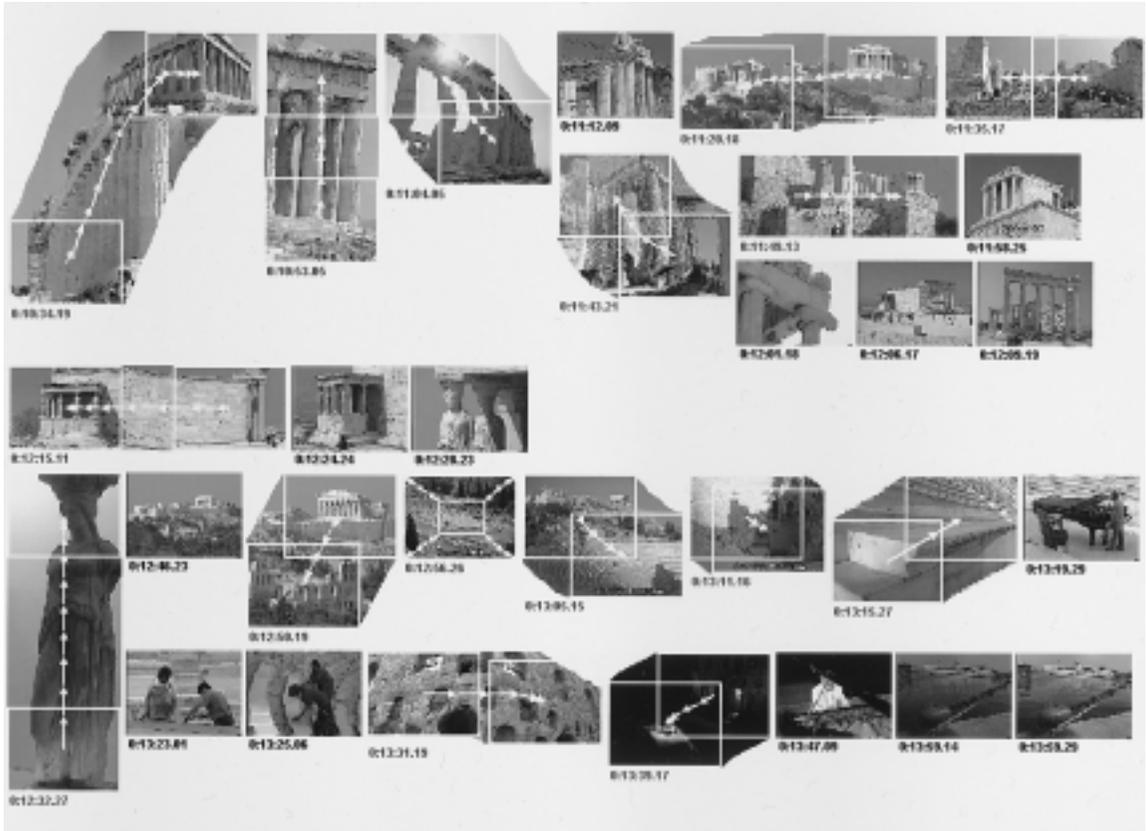


図8 PanoramaExcerpts インタフェース
Fig.8 PanoramaExcerpts interface.

が方法 A よりも効率的なレイアウトを出力したことを示す。ただし、提案法は最適解を与えるわけではないことに注意する。実際、方法 A よりも効率の悪いレイアウトを出力することもある^(注2)。

[実験 1] 一様乱数を用いて生成した 500 個の方形を 100 セット用意した。方形の高さ、幅は 0 以上 100 以下の値をとるように一様乱数によって決定した。ウィンドウ幅を 100 から 600 まで変化させながらパッキング率を測定した結果を図 7(a) に示す。 x 軸はウィンドウ幅を方形の幅の平均値 (50) で割った値であり、 y 軸はパッキング率である。これによれば、すべてのデータセットについて提案法は方法 A よりも効率が良いことがわかった。

[実験 2] 実際に映像から抽出されたデータ (80×60 画素の代表フレームアイコン 131 個、パノラマアイコン 47 個) に対してパッキング率を測定した。その結果を図 7(b) に示す。ウィンドウ幅が大きくなるに従っ

てパッキング率が 0.85 に近づいた。この結果は、提案法により方法 A に対して約 15% の画面スペースが節約され一覧性が向上したことを示す。

6. PanoramaExcerpts システムに関する考察

6.1 PanoramaExcerpts インタフェース

ギリシャに関する観光案内ビデオを処理して得られた PanoramaExcerpts インタフェースを図 8 に示す。PanoramaExcerpts インタフェースは以下の特長をもつ：

- パノラマアイコンによってショットの全体像がコンパクトに表現されている、
- パノラマアイコンの形状と軌跡表現によってカメラ操作の種別、方向、速度が直感的に表現されて

(注2): 例えば、幅 $W = 3$ のウィンドウに二つの方形 R_1 (幅 2, 高さ 1), R_2 (幅 1, 高さ 2) をレイアウトする場合。

いる、

- アイコンを時間順に一覧表示したことで映像の全体的な構成（コンテキスト）が表現されている。

PanoramaExcerpts インタフェースの有効性は対象とする映像の種別に依存する。PanoramaExcerpts インタフェースはカメラ操作を多く含んだ映像に対して特に効果がある。実験に使用した観光案内ビデオやドラマにはカメラ操作が比較的多く含まれていた。実験に使用した観光案内ビデオ（15分、180ショット）に含まれるショットの23.8%にあたる43個がパン、チルトを含んでおり、そのうち41個が正しく検出され、パノラマイコンとして表現された。しかしCM、ニュースなどにはカメラ操作がまれであったので、PanoramaExcerpts は従来の代表フレームのみの一覧とほぼ同等であった。

アイコンの時間順は、ほとんどの場合、無理なく追うことができたが、パノラマイコンの大きさのばらつきが大きくなると混乱を生じる可能性がある。

現在のプロトタイプでは、アイコンをクリックすることによりビデオを再生したり、アイコンのレイアウトを変更したり、不要なアイコンを削除したりといった簡単な編集をサポートしている。

カメラ操作軌跡をパノラマイコンに重畳して表示しているが、これは映像制作者の意図表現のツールである“絵コンテ”を映像から再構築しようとする試みの第一歩と位置づけることができる。絵コンテは映像の要点を静止画表現した映像要約の一つの理想形であり、これが映像から自動的に再構築できれば、映像の内容把握のためのツールとして活用できる。

6.2 処理速度

Motion JPEG方式で圧縮された、画像サイズ160×120画素の映像を毎秒10フレームに間引いてPC（Pentium II 300MHz）で処理した場合、入力映像の時間長の2/3程度で処理できた。したがって、リアルタイムに入力されるカメラ映像や放送映像も処理でき、リアルタイム映像のインデクシングなどへの応用も可能である。

レイアウト処理に要する計算時間は、画像再描画に要する時間と比べて無視できるほど小さかった。

7. むすび

本論文は、PanoramaExcerpts インタフェースを自動生成するために、映像を解析しパノラマイコンを抽出する方法、アイコンレイアウト方法について

提案した。実験を通して、PanoramaExcerpts インタフェースの有効性と、提案システムが実用的な検出性能、処理速度をもつことを示した。

今後の課題として、より効率的な映像要約のために、映像をより細かく分析、分類しより効果的な代表画像を抽出する方法を開発することが挙げられる。例えば、被写体の動きを考慮し、より効果的なアイコンを生成することも一つの課題である。

謝辞 本研究を進めるにあたり、有益な御助言をいただいたNTTヒューマンインタフェース研究所映像処理研究部徳永幸生前部長、御討論いただいた佐藤隆氏、南憲一氏、新倉康臣氏をはじめとするグループの皆様へ感謝致します。図1, 3, 4, 8に示した映像を文芸春秋社のご厚意により使用させていただきましたことを感謝致します。

文 献

- [1] P. Aigrain and P. Joly, “The automatic real-time analysis of film editing and transition effects and its applications,” *Computer and Graphics*, vol.18, no.1, pp.93–103, 1994.
- [2] P. Aigrain, H.-J. Zhang, and D. Petkovic, “Content-based representation and retrieval of visual media: A state-of-the-art review,” *Multimedia Tools and Applications*, vol.3, pp.179–202, 1996.
- [3] A. Akutsu and Y. Tonomura, “Video tomography: An efficient method for camerawork extraction and motion analysis,” *Proc. ACM Multimedia ’94*, ACM Press, pp.349–356, 1994.
- [4] P. Anandan, M. Irani, R. Kumar, and J. Bergen, “Video as an image data source: Efficient representations and applications,” *Proc. IEEE International Conference on Image Processing 95*, IEEE Computer Society Press, pp.318–321, 1995.
- [5] H. Aoki, S. Shimotuji, and O. Hori, “A shot classification method of selecting effective key-frames for video browsing,” *Proc. ACM Multimedia ’96*, ACM Press, pp.1–10, 1996.
- [6] K.A. Dowsland and W.B. Dowsland, “Packing Problems,” *European Journal of Operational Research*, vol.56, pp.2–14, 1992.
- [7] M. Massey and W. Bender, “Salient stills: Process and practice,” *IBM System Journal*, vol.35, no.3 and 4, pp.557–573, 1996.
- [8] M. Mills, J. Cohen, and Y.Y. Wong, “A magnifier tool for video data,” *Proc. of CHI ’92*, pp.93–98, 1992.
- [9] S. Mann and R. W. Picard, “‘Video orbits’: Characterizing the coordinate transformation between two images using the projective group,” MIT Media Laboratory Perceptual Computing Section Technical Report, no.278.
- [10] 長坂晃朗, 田中 譲, “カラービデオ映像における自動索引

付け法と物体検索法” 情処学論, vol.33, no.4, pp.543-550, 1992.

- [11] A. Pentland, R. Picard, G. Davenport, and K. Hasse, “Video and image semantics: Advanced tools for tele-communications,” *IEEE Multimedia*, vol.1, no.2, pp.73-75, 1994.
- [12] H. S. Sawhney and S. Ayer, “Compact representation of videos through dominant and multiple motion estimation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.18, no.8, pp.814-830, 1996.
- [13] M.A. Smith and T. Kanade, “Video skimming and characterization through the combination of image and language understanding techniques,” *Carnegie Mellon University Technical Report, CMU-CS-97-111*, 1997.
- [14] R. Szeliski, “Video mosaics for virtual environments,” *Computer Graphics & Applications*, pp.22-33, March 1996.
- [15] 谷口行信, 外村佳伸, 浜田 洋, “映像ショット切換え検出法とその映像アクセスインタフェースへの応用” 信学論 (D-II), vol.J79-D-II, no.4, pp.538-546, April 1996.
- [16] Y. Taniguchi, A. Akutsu, and Y. Tonomura, “PanoramaExcerpts: Extracting and packing panoramas for video browsing,” *Proc. ACM Multimedia 97*, Addison-Wesley, pp.427-436, 1997.
- [17] L. Teodosio and W. Bender, “Salient video still: Content and context preserved,” *Proc. ACM Multimedia '93*, pp.39-46, 1993.
- [18] L. Teodosio and M. Mills, “Panoramic overviews for navigating real-world scenes,” *Proc. ACM Multimedia '93*, pp.359-364, 1993.
- [19] Y. Tonomura and S. Abe, “Content oriented visual interface using video icons for visual database systems,” *Journal of Visual Languages and Computing*, vol.1, pp.183-198, 1990.
- [20] Y. Tonomura, A. Akutsu, K. Otsuji, and T. Sadakata, “Videomap and videospaceicon for anatomizing video content,” *Proc. INTERCHI '93*, IOS Press, pp.131-138, 1993.
- [21] 上田博唯, 宮武孝文, 吉沢 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.
- [22] M.M. Yeung and B.-L. Yeo, “Video visualization for compact presentation and fast browsing of pictorial content,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol.7, no.5, pp.771-785, 1997.
- [23] H.J. Zhang, C.Y. Low, S.W. Smoliar, and J.H. Wu, “Video parsing, retrieval and browsing: An integrated and content-based solution,” *Proc. ACM Multimedia '95*, ACM Press, pp.15-24, 1995.

(平成10年5月21日受付, 9月3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. 映像処理の研究に従事. 現在, NTT サイバースペース研究所研究主任. 平7 本学会学術奨励賞, ACM Multimedia'97 Best Paper Award 受賞. 情報処理学会, ACM 各会員.



阿久津明人 (正員)

昭63 千葉大・工卒. 平2 同大大学院修士課程了. 同年, 日本電信電話 株 入社. NTT サイバースペーション研究所にて映像ハンドリング技術構築を目的とした映像処理の研究に従事. 平5 本学会学術奨励賞受賞. 現在, NTT サイバースペーション研究所研究主任. 映像情報メディア学会, ACM 各会員.



外村 佳伸 (正員)

昭54 京大・工・電子卒. 昭56年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 サイバースペーション研究所主幹研究員. 映像情報メディア学会, IEEE, ACM 各会員.