

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第4972095号  
(P4972095)

(45) 発行日 平成24年7月11日(2012.7.11)

(24) 登録日 平成24年4月13日(2012.4.13)

(51) Int. Cl. F I  
**G06T 3/00 (2006.01)** G06T 3/00 300  
**G06T 7/20 (2006.01)** G06T 7/20 C

請求項の数 14 (全 25 頁)

(21) 出願番号	特願2008-539616 (P2008-539616)	(73) 特許権者	508139491
(86) (22) 出願日	平成18年11月15日 (2006.11.15)		イッサム リサーチ ディベロップメント
(65) 公表番号	特表2009-516257 (P2009-516257A)		カンパニー オブ ザ ヘプライ ユニ
(43) 公表日	平成21年4月16日 (2009.4.16)		バーシティー オブ エルサレム
(86) 国際出願番号	PCT/IL2006/001320		イスラエル国 エルサレム 91390,
(87) 国際公開番号	W02007/057893		ピー. オー. ボックス 39135, キバ
(87) 国際公開日	平成19年5月24日 (2007.5.24)		トラム, エドモンドサフラキャンパス, ハ
審査請求日	平成21年11月13日 (2009.11.13)		イテックパーク
(31) 優先権主張番号	60/736, 313	(74) 代理人	100096024
(32) 優先日	平成17年11月15日 (2005.11.15)		弁理士 柏原 三枝子
(33) 優先権主張国	米国 (US)	(74) 代理人	100125520
(31) 優先権主張番号	60/759, 044		弁理士 高橋 剛一
(32) 優先日	平成18年1月17日 (2006.1.17)	(74) 代理人	100155310
(33) 優先権主張国	米国 (US)		弁理士 柴田 雅仁

最終頁に続く

(54) 【発明の名称】 映像概要を生成する方法およびシステム

(57) 【特許請求の範囲】

【請求項 1】

ビデオカメラによって取得された第1の動的シーンの映像フレームのソースシーケンスを、第2の動的シーンを表示する映像フレームのより短い概要シーケンスに変換することによって、映像概要 (video synopsis) を生成するための方法であって、当該方法が、

少なくとも一つのオブジェクトの動作を示す前記ソースシーケンスの映像フレームのサブセットを取得するステップであって、各オブジェクトが、前記ソースシーケンスの少なくとも3の異なるフレームからの、各フレーム内で互いに連結されたピクセルのサブセットであるステップを具え、当該方法が、

前記ソースシーケンスから少なくとも3のソースオブジェクトを選択し、時間サンプリング (temporal sampling) によって、各々の選択されたソースオブジェクトから1又はそれ以上の概要オブジェクトをサンプリングするステップであって、前記時間サンプリングが、ソースシーケンスのN個のフレームを映像シーケンスのM個 (NはMとは異なる) のフレームにマッピングするものであり、前記選択が、オブジェクトの現れる時間、オブジェクトの種類、およびオブジェクトの動作の種類のうちの少なくとも1つを含む選択基準に基づくものであるステップと、

各概要オブジェクトについて前記映像概要における表示を開始する各表示時間を決定するステップであって、映像概要の全体の再生時間が前記ソースシーケンスの再生時間よりも短くなるように、概要オブジェクトの各々について表示時間を決定するステップと、

10

20

前記ソースシーケンスのそれぞれ異なる時間で得られた少なくとも2の概要オブジェクトが、前記映像概要中に同時に表示されるように、前記第1の動的シーンの前記オブジェクトの空間的な位置を変更することなく、概要オブジェクトをそれぞれの所定の表示時間で表示することによって前記映像概要を生成するステップと、によって特徴付けられる方法。

【請求項2】

請求項1に記載の方法において、前記オブジェクトのうちの1つが背景オブジェクトであることを特徴とする方法。

【請求項3】

請求項2に記載の方法において、前記オブジェクトと前記背景とを継ぎ目のない映像に  
つなぎ合わせることを特徴とする方法。 10

【請求項4】

請求項1乃至3のいずれか1項に記載の方法において、前記ソースオブジェクトが選択され、各概要オブジェクトの表示を開始する各時間が費用関数を最適化するように決定され、前記費用関数が、前記映像概要の長さとその質的な映像基準 ( qualitative video metrics ) との間の妥協点を決定するものであることを特徴とする方法。

【請求項5】

請求項1乃至4のいずれか1項に記載の方法において、前記ソースシーケンスが固定位置で軸に対し回転するカメラによって取得され、前記概要シーケンスにコピーする前に少なくとも前記ソースオブジェクトの少なくとも2つを空間的に移動させることを特徴とする方法。 20

【請求項6】

請求項1乃至4のいずれか1項に記載の方法において、前記ソースシーケンスが固定位置で静止カメラによって取得されることを特徴とする方法。

【請求項7】

請求項1乃至6のいずれか1項に記載の方法において、少なくとも3の異なるソースフレームが時間的に連続していることを特徴とする方法。

【請求項8】

請求項1乃至7のいずれか1項に記載の方法において、前記ソース映像シーケンスにおいて同時に発生する2のイベントが、前記概要シーケンスにおいて異なる時間に表示されることを特徴とする方法。 30

【請求項9】

請求項1乃至8のいずれか1項に記載の方法が、監視用映像概要、動画の活動密度 ( activity density ) の増加、映像インデックスのいずれかに用いられることを特徴とする方法。

【請求項10】

請求項9に記載の方法が、前記概要シーケンス内の各ピクセルについて、前記ソースシーケンス内の対応するピクセルへのポインタを維持するステップを含むことを特徴とする方法。 40

【請求項11】

請求項1乃至10のいずれか1項に記載の方法であって、

( a ) 前記ソースシーケンスのフレーム間の画像動作パラメータを算出し、

( b ) 前記第1の動的シーン内の静止したオブジェクトが整列されたソースシーケンス内で静止するように、前記ソースシーケンス内の映像フレームを移動させることにより、前記整列されたソースシーケンスを与えるように、当該ソースシーケンスを事前に整列させるステップを含むことを特徴とする方法。

【請求項12】

第1の動的シーンの映像フレームのソースシーケンスを、第2の動的シーンを表示する少なくとも2つの映像フレームの概要シーケンスに変換するシステム ( 10 ) であって、 50

当該システムが、

少なくとも一つのオブジェクトの動作を示す前記ソースシーケンス内の映像フレームのサブセットを格納し、各々のオブジェクトが、少なくとも3の異なるソースフレームからの、各フレーム内で互いに連結されたピクセルのサブセットであり、前記時間サンプリングが、ソースシーケンスのN個のフレームを映像シーケンスのM個のフレーム(NはMとは異なる)にマッピングするものであり、前記選択が、オブジェクトの現れる時間、オブジェクトの種類、およびオブジェクトの動作の種類のうち少なくとも一つを含む選択基準に基づくものである、第1のメモリ(12)を具え、前記システムが、

前記ソースシーケンスから少なくとも3のオブジェクトを選択し、時間サンプリングによって、各々の選択されたソースオブジェクトから1又はそれ以上の概要オブジェクトをサンプリングする、前記第1のメモリ(12)に連結された選択ユニット(13)と、

10

各概要オブジェクトについて、映像概要における表示を開始するために各表示時間を決定し、前記ソースシーケンスのそれぞれ異なる時間で得た少なくとも2の概要オブジェクトが、前記映像概要中に同時に表示されるように、前記オブジェクトの空間的位置又は前記第1の動的シーンにおいてそこから得られる前記それぞれのオブジェクトを変更することなく、概要オブジェクト、又は、各所定の表示時間でそこから得られたオブジェクトとを表示することによって前記映像概要を生成する、フレーム生成部であって、映像概要の全体の再生時間が前記ソースシーケンスの再生時間よりも短くなるように、概要オブジェクトの各々について表示時間を決定するフレーム生成部(14)と、

前記フレーム生成部に連結されて、前記概要シーケンスのフレームを蓄積する、第2のメモリ(15)と、

20

前記第2の動的シーンを表示すべく、前記第2のメモリ(15)にディスプレイ装置(16)を連結する手段とによって特徴付けられるシステム。

【請求項13】

請求項12に記載のシステムにおいて、前記フレーム生成部(14)が、前記概要シーケンスにコピーする前に、前記ソースオブジェクトの少なくとも2つを空間的に移動させる移動ユニット(17)を具えることを特徴とするシステム。

【請求項14】

コンピュータプログラムであって、当該プログラムがコンピュータ上で実行されるときに、請求項1乃至11のいずれか1項に記載の方法を実行するコンピュータプログラムコードを具えることを特徴とするコンピュータプログラム。

30

【発明の詳細な説明】

【技術分野】

【0001】

関連出願

本出願は、2006年11月2日に出願された米国第10/556,601(Peleg等)「Method and sytem for spatio-temporal video warping」の一部継続出願であり、2006年5月11日に公開されたWO2006/048875に対応し、さらに、2005年11月15日に出願された暫定出願第60/736,313および2006年1月17日に出願された60/759,044の優先権を主張し、これらの出願の総ての内容は、参照によりここに組み込まれている。

40

【0002】

本発明は、一般に、新しい画像および映像が、シーンの複数のオリジナル映像から部分を統合することにより生成されるレンダリングに基づく画像および映像に関連する。特に、本発明は、映像の抽象化あるいは概要(synopsis)を目的とするこのような技術に関連する。

【0003】

従来技術

本発明の背景として関連があると考えられる従来技術の引例が、以下に記載されており、これらの内容は、ここに組み込まれている。追加の引例が、前述した米国暫定出願第6

50

0 / 7 3 6 , 3 1 3 および 6 0 / 7 5 9 , 0 4 4 に記載されており、これらの内容は、参照によりここに組み込まれている。ここに記載された引用の確認は、ここに開示された本発明の特許性に関連する方法であることを意味するものではない。各引用は、角括弧に囲まれた数字により特定され、したがって、従来技術は、明細書全体を通して、角括弧に囲まれた数字により引用される。

[ 1 ] A. Agarwala, M. Dontcheva, M. Agrawala, S. Drucker, A. Colburn, B. Curless, D. Salesin, and M. Cohen. Interactive digital photomontage. 2004年のシーグラフ (SIGGRAPH) の 294 から 302 ページ。

[ 2 ] A. Agarwala, K. C. Zheng, C. Pal, M. Agrawala, M. Cohen, B. Curless, D. Salesin, and R. Szeliski. Panoramic video textures. 2005年のシーグラフの 821 から 827 ページ。 10

[ 3 ] J. Assa, Y. Caspi, and D. Cohen-Or. Action synopsis: Pose selection and illustration. 2005年のシーグラフの 667 から 676 ページ。

[ 4 ] O. Boiman and M. Irani. Detecting irregularities in images and in video. 2005年北京における ICCV のページ I : 462 から 469 。

[ 5 ] A. M. Ferman and A. M. Tekalp. Multiscale content extraction and representation for video indexing. Proc. of SPIE, 3229:23-31, 1997.

[ 6 ] M. Irani, P. Anandan, J. Bergen, R. Kumar, and S. Hsu. Efficient representations of video sequences and their applications. Signal Processing: Image Communication, 8(4):327-351, 1996. 20

[ 7 ] C. Kim and J. Hwang. An integrated scheme for object-based video abstraction. 2000年ニューヨークにおける ACM マルチメディア (ACM Multimedia) の 303 - 311 。

[ 8 ] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. Science, 4598(13):671-680, 1983.

[ 9 ] V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts? 2002年 ECCV のページ 65 から 81 。

[ 10 ] Y. Li, T. Zhang, and D. Treter. An overview of video abstraction techniques. Technical Report HPL-2001-191, HP Laboratory, 2001.

[ 11 ] J. Oh, Q. Wen, J. Lee, and S. Hwang. Video abstraction. In S. Deb, editor, Video Data Management and Information Retrieval, pages 321-346. Idea Group Inc. and IRM Press, 2004. 30

[ 12 ] C. Pal and N. Jojic. Interactive montages of sprites for indexing and summarizing security video. In Video Proceedings of CVPR05, page 11: 1192, 2005.

[ 13 ] A. Pope, R. Kumar, H. Sawhney, and C. Wan. Video abstraction: Summarizing video content for retrieval and visualization. In Signals, Systems and Computers, pages 915-919, 1998.

[ 14 ] WO2006/048875 Method and system for spatio-temporal video warping, pub. May 11, 2006 by S. Peleg, A. Rav-Acha and D. Lischinski. これは、2005年11月2日に出版された米国第 10 / 556 , 601 に対応する。 40

[ 15 ] A. M. Smith and T. Kanade. Video skimming and characterization through the combination of image and language understanding. 1998年の CATVD の 61 から 70 ページ。

[ 16 ] A. Stefanidis, P. Partsinevelos, P. Agouris, and P. Doucette. Summarizing video datasets in the spatiotemporal domain. 2000年の DEXA のワークショップの 906 から 912 。

[ 17 ] H. Zhong, J. Shi, and M. Visontai. Detecting unusual activity in video. 2004年の CVPR のページ 819 から 826 。

[ 18 ] X. Zhu, X. Wu, J. Fan, A. K. Elmagarmid, and W. G. Aref. Exploring video content structure for hierarchical summarization. Multimedia Syst, 10(2): 98-1 50

15, 2004.

[ 1 9 ] J. Barron, D. Fleet, S. Beauchemin and T. Burkitt.. Performance of optical flow techniques, volume 92, pages 236-242.

[ 2 0 ] V. Kwatra, A. Sch d l, I. Essa, G. Turk and A. Bobick. Graphcut texture s: image and video synthesis using graph cuts. 2 0 0 3年のシーグラフの2 2 7から2 8 6ページ。

[ 2 1 ] C. Kim and J. Hwang, Fast and Automatic Video Object Segmentation and Tracking for Content-Based Applications, IEEE Transactions on Circuits and Systems for Video Technology, Vol. 12, No. 2, February 2002, pp 122-129.

[ 2 2 ] 米国特許第 6 , 6 6 5 , 0 0 3

10

【背景技術】

【 0 0 0 4 】

映像の概要 ( video synopsis ) は、映像のブラウジングおよび検索を可能にする一時的なコンパクト表現である。

【 0 0 0 5 】

映像の概要のための2つの主要なアプローチがある。一方のアプローチでは、一組の特徴的な画像 ( キーフレーム ) が、オリジナルの映像シーケンスから選択される。選択されたキーフレームは、映像を最もよく表わすフレームである [ 7 , 1 8 ]。他方のアプローチでは、短い映像シーケンスの収集が選択される [ 1 5 ]。第2のアプローチは、コンパクトではないが、シーンの動的性 ( dynamics ) に良い影響を与える。これらのアプローチ ( および他のアプローチ ) は、映像の抽象化の広範囲な調査に記されている [ 1 0 , 1 1 ]。

20

【 0 0 0 6 】

前述した双方のアプローチでは、総てのフレームが、基礎的な基本的要素として利用される。別の方法は、映像のインデックスのためにメタデータと共にモザイク画像を利用する [ 6 , 1 3 , 1 2 ]。この方法では、静的な概要画像 ( synopsis image ) は、時間の異なるオブジェクトを含む。

【 0 0 0 7 】

オブジェクトに基づくアプローチが知られており、このアプローチでは、オブジェクトが入力映像から抽出される [ 7 , 5 , 1 6 ]。しかしながら、これらの方法は、重要なキーフレームを特定するオブジェクトの検出を利用し、異なる時間間隔の活動を組み合わせない。

30

【 0 0 0 8 】

繰り返される最小のカット ( iterated min-cut ) を利用して単一のパノラマ画像を生成し [ 1 ]、繰り返される最小のカットを利用してパノラマ映像を生成する [ 2 ] する方法が、本分野で知られている。この2つの方法では、 ( 入力フレームの数の ) 指数の複雑さに関する問題が近似され、したがって、これらは、少数のフレームにより適切である。本分野の関連する研究は、最小のカットを利用して2つの映像を組み合わせることに関連する [ 2 0 ]。

【 0 0 0 9 】

40

WO 2 0 0 6 / 0 4 8 8 7 5 [ 1 4 ] は、映像の一時的な流れを操作する方法およびシステムを開示している。第1の動的シーンの映像フレームの第1のシーケンスは、第2の動的シーンを表示する映像フレームの第2のシーケンスに変換され、一の態様では、第1の動的シーン内の少なくとも一つの特徴の場合、映像フレームの第1のシーケンスの各部分は、映像フレームの第1のシーケンスの周辺部分よりも異なるレートで標本化され、標本化された部分は、第2のシーケンスの対応するフレームにコピーされる。これにより、動的シーン内の特徴の一時的な同期が変更できる。

【発明の開示】

【課題を解決するための手段】

【 0 0 1 0 】

50

本発明の第1の態様によると、第1の動的シーンの映像フレームの第1のシーケンスを、第2の動的シーンを表示する少なくとも2つの映像フレームの第2のシーケンスに変換するコンピュータが実施する方法が提供され、当該方法が、

(a)  $x$ ,  $y$  座標それぞれに配置された複数のピクセルを備える少なくとも一つのオブジェクトの動作を示す前記第1のシーケンスの映像フレームのサブセットを取得するステップと、

(b) 各映像フレームで前記第1の動的シーン内の少なくとも一つのオブジェクトの空間的に重複しない外観 (non-spatially overlapping appearance) を示す前記サブセット部分から選択するステップと、

(c) 前記オブジェクトのピクセルの  $x$ ,  $y$  座標それぞれを変更することなく、前記部分を、少なくとも3つの異なる入力フレームから前記第2のシーケンスの少なくとも2つの連続したフレームにコピーして、前記第2のシーケンスのうちの少なくとも一つのフレームが、前記第1のシーケンスの異なるフレームで現れる少なくとも2つの部分を含むようにするステップとを備える。

#### 【0011】

本発明の第2の態様によると、第1の動的シーンの映像フレームの第1のシーケンスを、第2の動的シーンを表示する少なくとも2つの映像フレームの第2のシーケンスに変換するシステムが提供され、当該システムが、

$x$ ,  $y$  座標に配置された複数のピクセルを備える少なくとも一つのオブジェクトの動作を示す前記第1のシーケンス内の映像フレームのサブセットを格納する第1のメモリと、

各映像フレームで、前記第1の動的シーン内の少なくとも一つのオブジェクトの空間的に重複しない外観を示す前記サブセット部分から選択するための、前記第1のメモリに連結される選択ユニットと、

前記オブジェクト内の  $x$ ,  $y$  座標それぞれを変更することなく、前記部分を、少なくとも3つの異なる入力フレームから、前記第2のシーケンスの少なくとも2つの連続したフレームにコピーし、前記第2のシーケンスのフレームのうちの少なくとも一つが、前記第1のシーケンス内の異なるフレームで現れる少なくとも2つの部分を含むフレーム生成部と、

前記第2のシーケンスのフレームを蓄積する第2のメモリとを備える。

#### 【0012】

本発明の第3の態様はさらに、動的シーンを表示する一連の出力映像フレームを確実に実施するデータキャリアを備えており、前記出力映像の少なくとも2つの連続したフレームは、 $x$ ,  $y$  座標それぞれを有する複数のピクセルであって、オブジェクト内のピクセルの  $x$ ,  $y$  座標それぞれを変更することなく、少なくとも3つの異なる入力フレームのオブジェクトの部分から生じる複数のピクセルを備え、前記出力映像フレームのうちの少なくとも一つが、異なる入力フレームで現れる少なくとも2つの部分を含む。

#### 【0013】

本発明により開示された動的な映像概要 (video synopsis) は、前述した従来の映像の抽象化手段と、以下の2つの特定において相違する。(i) 映像概要自体が、シーンの動的性を表示する映像である。(ii) 時空間の重複 (spatio-temporal redundancy) をできる限り低減すべく、活動の相対的なタイミングを変更する。

#### 【0014】

例えば、図1の空間時間量 (space-time volume) で表わされる概略的な映像クリップを検討する。映像は、地面を歩く人で始まり、活動のない期間の後、鳥が空を飛んでいる。活動のないフレームは、多くの映像の抽象化の方法では省略される。映像概要は、人と鳥を同時に表示することにより実質的にコンパクトである。これは、他の活動がこの空間的な位置で生じない場合、オリジナルの時間間隔から別の時間間隔にイベントをシフトさせることによる画像領域の最適利用を可能にする。このような操作は、初めに [14] に示したように、年代のコンシステンシー (chronological consistency) を緩和する

## 【 0 0 1 5 】

また、本発明は低レベルの方法を提供し、マルコフ確率場の最適化を利用して映像概要を生成する。

## 【 0 0 1 6 】

本発明により提供される選択肢のうちの一つは、単一のオブジェクトの複数の動的な外観を表示する能力である。この効果は、移動するオブジェクトの従来の映像概要で用いられる「ストロボ観察 (strobe-scopic)」画像の一般化である [ 6 , 1 ]。これを実行する2つの異なるスキームが提供される。第1のスキームでは、異なる時間の事例におけるオブジェクトのスナップショットは、開始位置から終了位置までの映像全体のオブジェクトの経過の指標を提供すべく出力映像で提供される。第2のスキームでは、オブジェクトは、開始あるいは終了位置を定義しないが、ランダムに予測不可能に移動する。この場合、異なる時間の事例におけるオブジェクトのスナップショットが再び、出力映像として提供されるが、この時間は、実際よりも多いオブジェクトの印象を与える。双方のスキームが共通点は、入力映像から異なる時間に取得される複数のスナップショットが、空間的な重複を避け、且つ関心のあるオブジェクトの動的な経過に寄与しないコピーしない方法により、入力映像から出力映像にコピーされることである。

## 【 0 0 1 7 】

本発明の内容および添付の請求項では、「映像 (video)」の語は、最も一般的な用語の「動画 (movie)」と同義であり、これは、ポスト処理に従うコンピュータ画像ファイルとしてアクセス可能であり、任意の種類動画ファイル、例えばデジタル、アナログを含む。カメラは、回転およびズームが可能な所定の位置にあることが好適であるが、これまで提案された技術で行われるような影響を受けやすい並進運動ではない。本発明に関連するシーンは、例えば、米国特許 6,665,003 [ 2 2 ] および動的なシーンを表示しない立体的な画像の表示を目的とする他の引例で処理される静的なシーンと反対に動的であり、連続したフレームが空間的および時間的な連続性を有する。本発明の一の多様によるうと、我々はこの問題を、グラフ上の最大の流れを見つけることにより多項式時間で説くことができる単一の最小のカットの問題として定式化する [ 5 ]。

## 【 0 0 1 8 】

本発明を説明するために、「空間時間量」と称する構成が利用され、動的なパノラマ映像を生成する。空間時間量は、総てのフレームを時間軸に沿って順次スタックすることにより画像の入力シーケンスから構成される。しかしながら、実際の実施に関する限り、例えば、動的なソースシーンの2Dフレームを時間内にスタックすることにより空間時間量を実際に構成する必要はないことを理解すべきである。さらに通常は、ソースフレームは、ターゲットフレームを構成するために個々に処理されるが、空間時間量を概念的な構成でなく物理的な構成であるかのように参照することは理解を助けるであろう。

## 【 0 0 1 9 】

本発明を理解し、本発明が実際にどのように実行されるか理解するために、好適な実施例は、限定ではなく添付図面を参照することにより説明される。

## 【 発明を実施するための最良の形態 】

## 【 0 0 2 0 】

## 1. 活動検出

本発明は、総ての入力ピクセルが「重要度」のレベルが付されていると仮定する。これ以降、活動レベルを「重要度」のレベルとして使用するが、所望の応用例に基づいて他の測定値を「重要度」として使用できることは明らかであろう。重要度 (あるいは活動) レベルの評価が想定され、それ自体は本発明の特徴ではない。これは、不規則性を検出し [ 4 , 1 7 ]、オブジェクト検出を移動させ、オブジェクトトラッキングを行う様々な手段のうちの一つを利用して行うことができる。代替的に、これは、顔認識などの認識アルゴリズムに基づいて行うことができる。

## 【 0 0 2 1 】

例えば、単一旦つ共通で使用される活動インジケータが選択され、入力ピクセル  $I(x$

10

20

30

40

50

$(x, y, t)$  は、位置  $(x, y)$  における時間的なメジアンからの色差が所定の閾値よりも大きい場合、「アクティブ」とされる。アクティブピクセルは、特有の関数により既定される。

$$\chi(p) = \begin{cases} 1 & \text{if } p \text{ is active} \\ 0 & \text{otherwise,} \end{cases}$$

【0022】

活動インジケータからノイズを除去すべく、メジアンフィルタが、概要 ( synopsis ) 処理を継続する前に適用される。

【0023】

連続的な活動測定を利用できるが、本発明は2元の事例に焦点を合わせる。連続的な活動の測定は、以下の詳細な説明のほぼ総ての等式を僅かに変更するだけで行うことができる [ 4, 17, 1 ]。

【0024】

我々は、映像概要の算出のための2つの異なる実施例を記載する。1つのアプローチ ( セクション2 ) は、グラフカットを用いたグラフ表現および費用関数の最適化を利用する。別のアプローチ ( セクション3 ) は、オブジェクトセグメンテーションおよびトラッキングを利用する。

【0025】

2. エネルギーの最小化による映像概要

入力映像シーケンスのN個のフレームを3Dの空間時間量  $I(x, y, z)$  で表示させ、この場合、 $(x, y)$  はこのピクセルの空間的な座標であり、 $1 \leq t \leq N$  がフレームの数である。

【0026】

我々は、以下の特性を有する映像概要  $S(x, y, z)$  を生成する。

- ・映像概要  $S$  は、オリジナル映像  $I$  よりも実質的に短くあるべきである。
- ・オリジナル映像の最大の「活動」は、要約映像内で現れるべきである。
- ・映像概要内のオブジェクトの動作は、オリジナル映像内の動作と同様であるべきである。
- ・映像概要は、見栄えがよく、目に見える継ぎ目 ( シーム ) あるいは断片化したオブジェクトは阻止されるべきである。

【0027】

前述した特性を有する映像概要  $S$  は、マッピング  $M$  を利用して生成され、概要  $S$  内の総ての座標  $(x, y, z)$  に、 $I$  からのソースピクセルの座標を割り当てる。我々は、空間的な位置を固定しつつ、ピクセルのタイムシフトに焦点を合わせる。したがって、概要ピクセル  $S(x, y, z)$  は、入力ピクセル  $I(x, y, M(x, y, t))$  から生じる。タイムシフト  $M$  は、エネルギーの最小化の問題を解決することにより得られ、費用関数は、以下の等式により得られる。

$$E(M) = E_a(M) + \alpha E_d(M), \quad (1)$$

ここで、 $E_a(M)$  は活動の損失を示し、 $E_d(M)$  は継ぎ目における不連続性を示す。活動の損失は、映像概要  $S$  内で現れない入力映像  $I$  のアクティブピクセルの数であろう。

$$E_a(M) = \sum_{(x,y,t) \in I} \chi(x,y,t) - \sum_{(x,y,t) \in S} \chi(x,y,M(x,y,t)). \quad (2)$$

【0028】

不連続性費用 ( discontinuity cost )  $E_d$  は、映像概要内の空間および時間的な隣接部

10

20

30

40

50

分と、入力映像内の対応する隣接部分との間の継ぎ目における色差の合計として既定される（同様の等式が[1]に存在する）。

$$E_d(M) = \sum_{(x,y,t) \in S} \sum_i \|S((x,y,t) + e_i) - I((x,y,M(x,y,t)) + e_i)\|^2 \quad (3)$$

ここで、 $e_i$  は、6つの空間および時間的な隣接部分を表わす6つのユニットである。

【0029】

図2aおよび2bは、費用関数を最小化することにより短い映像概要を生成する空間および時間的な処理を表わす概略図であり、移動するオブジェクトの動作は、図中の「活動ストリップ」により表わされる。上図はオリジナルの映像を示しており、下図は映像概要を示している。特に図2aでは、短い映像概要Sは、多くのアクティブピクセルを含むことにより入力映像Iから生成される。円滑にすべく、S内のピクセルAがI内のピクセルBに対応する場合、「境界を越えた」隣接部分は同様であるべきである。最適なM最小化(3)を発見することは、非常に大きな最適化の問題である。近似解が図2bに示されており、映像概要内の連続したピクセルは、連続した入力ピクセルから生成されるように限定されている。

10

【0030】

費用関数E(M)(等式1)は3次元のマルコフ確率場に対応し、各ノードは、出力映像の3次元体積内のピクセルに対応し、入力フレームに対応する時間的価値が割り当てることができることに留意する。ノードの重みは、活動費用により決定され、ノード間のエッジは、不連続性費用に応じて決定される。したがって、費用関数は、反復するグラフカット(iterative graph-cuts)などのアルゴリズムにより最小化することができる[9]

20

【0031】

#### 2.1.2 Dグラフを利用した限定的な解決方法

映像概要内の各ピクセルが任意の時間に生じることを可能にする等式(1)の最適化は、大きな問題である。例えば、5秒の映像概要に要約される3分の入力映像は、それぞれが5400ラベルを有する約 $2^{25}$ ノードを有するグラフを生じる。

【0032】

水平な経路を移動する動的なテクスチャあるいはオブジェクトの場合、3次元のMRFは、この問題を1次元の問題に低減させることにより効果的に解決できることが[2]に記載されていた。この研究において、我々は、より一般的な方法で移動するオブジェクトに取り組み、したがって、我々は異なる制約を用いる。映像概要S内の連続したピクセルは、入力映像I内の連続したピクセルから生じるように制限されている。この制限の下では、3次元のグラフは、2次元のグラフに減らされ、各ノードは、概要動画内の空間的な位置に対応する。各ノードM(x,y)のラベルは、図2bに示すように、Sの第1のフレームに示すI内のフレーム数tを決定する。M(x<sub>1</sub>,y<sub>1</sub>) M(x<sub>2</sub>,y<sub>2</sub>)の場合、継ぎ目は、2つの隣接する位置(x<sub>1</sub>,y<sub>1</sub>)と(x<sub>2</sub>,y<sub>2</sub>)との間に存在し、継ぎ目に沿った不連続性費用E<sub>d</sub>(M)は、S内の総てのフレームに跨るこの空間的な位置における色差の合計である。

30

40

$$E_d(M) = \sum_{x,y} \sum_i \sum_{t=1}^K \|S((x,y,t) + e_i) - I((x,y,M(x,y,t)) + e_i)\|^2 \quad (4)$$

ここで、 $e_i$  は、4つの空間的な隣接部分を示す4つの単位ベクトルである。

【0033】

各ノードのラベルの数はN-Kであり、NおよびKは、入力および出力映像それぞれのフレームの数である。各ピクセルの活動の損失は、

50

$$E_a(M) = \sum_{x,y} \left( \sum_{t=1}^N \chi(x,y,t) - \sum_{t=1}^K \chi(x,y, M(x,y) + t) \right).$$

である。

【 0 0 3 4 】

### 3 . オブジェクトに基づく概要

前述した動的な映像概要の低いレベルのアプローチは、目に見える継ぎ目を防止するといった局所的な特性に限定されている。高いレベルのオブジェクトに基づく特性は、オブジェクトが検出できる場合に取り入れることができる。例えば、ストロボ効果を避けることは、体積内の各オブジェクトの検出およびトラッキングを必要とする。このセクションでは、動的な映像概要のためのオブジェクトに基づくアプローチの実施について説明する。いくつかのオブジェクトに基づく映像の要約方法は、文献（例えば、[ 7 , 5 , 1 6 ]）に存在し、これらは総て、重要なフレームの選択するために、検出されたオブジェクトを利用する。これらの方法とは違い、本発明は、時間内にオブジェクトを移動させ、入力シーケンス内に現れない新しい概要フレームを生成し、空間および時間を有効に利用する。

10

【 0 0 3 5 】

一実施例では、移動するオブジェクトは、前述したように、各ピクセルを時間的なメジアンと比較し、この違いを閾値化することにより検出される。この後、空間的なメジアンフィルタを利用したノイズクリーニングが行われ、空間および時間的に関連する構成要素が組み合わされる。文献内には、このタスクのために利用されるオブジェクト検出およびトラッキングのための多くの他の方法が存在することは理解できるであろう（例えば、[ 7 , 1 7 , 2 1 ]）。オブジェクト検出およびトラッキングの各処理は、一組のオブジェクトを生成し、各オブジェクト  $b$  は、特有の関数により表わされる。

20

$$\chi_b(x,y,t) = \begin{cases} 1 & \text{if } (x,y,t) \in b \\ 0 & \text{otherwise,} \end{cases} \quad (5)$$

30

【 0 0 3 6 】

図 3 a , 3 b , 3 c は、本発明による時間的な再配置の例を示した図である。各図の上図は、オリジナルの映像を表わしており、下図は、映像概要を表わしており、移動するオブジェクトの動作は、図中の「活動ストリップ」により表わされる。図 3 a は、異なる時間に記録され、映像概要の同じ時間間隔で移動する 2 つのオブジェクトを示している。図 3 b は、長い期間移動し、短い時間間隔を有するセグメントに分割される単一のオブジェクトを示しており、これは、同時に表示され、動的なストロボ効果を生成する。図 3 c は、オブジェクトの体積がセグメントに分割される場合に、オブジェクトの交差が概要を妨げないことを示している。

40

【 0 0 3 7 】

各オブジェクトから、セグメントは、オブジェクトが現れるフレームのサブセットを選択することにより生成される。このようなセグメントは、異なる時間間隔を表わすことができ、任意で異なる標本化レートで取得される。

【 0 0 3 8 】

映像概要  $S$  は、以下の処理を用いて、入力映像  $I$  から構成される。

( 1 ) オブジェクト  $b_1 \dots b_r$  は、入力映像  $I$  から抽出される。

( 2 ) 一組の重ならないセグメント  $B$  は、オリジナルのオブジェクトから選択される。

( 3 ) 時間的なシフト  $M$  は、選択された各セグメントに適用され、短い映像概要を生成するとともに、オブジェクト間の重なりを阻止し、継ぎ目のないステッチ ( stitching )

50

を可能にする。これは、図 1 および図 3 a 乃至 3 c に説明されている。図 4 は、映像概要の単一のフレームが図 3 b に示す動的なストロボ効果を利用する例を示した図である。

【 0 0 3 9 】

我々が処理 ( 2 ) および ( 3 ) を時間内に選択およびシフトさせ、短くて継ぎ目の無い映像概要を取得するため、処理 ( 2 ) および ( 3 ) は関連している。前述した処理 ( 2 ) および ( 3 ) は、完全である必要はないことは理解できるであろう。「空間的に重複しないセグメント (non-spatially overlapping segment)」という場合には、小さな重複部分が許容され、「重なりを阻止する」という場合には、時間内に移動するオブジェクト間の小さな重なりは許容されるが、これは、視覚的に魅力的な映像を得るために最小限にすべきである。

10

【 0 0 4 0 】

オブジェクトに基づく表示では、生成された概要内のピクセルは、複数のソース (異なるオブジェクトから生成され) を有し、したがって、我々は、総てのオブジェクトが統合される後の処理ステップを付加する。背景画像は、シーケンスの総てのフレームにおけるピクセルのメジアン値を取得することにより生成される。次に、選択されたオブジェクトは、各フレーム内のピクセルとメジアン画像との間の距離 (RGB 空間内) に比例した重み付けを利用して融合可能である。このステッチ機構は、[ 6 ] で使用されるものと同様である。

【 0 0 4 1 】

我々は、単一の概要ピクセルにマップされる一組の総てのピクセル  $(x, y, t) \in S$  を、 $src(x, y, t)$  として規定し、我々は、オブジェクト (あるいはセグメント)  $b$  内の (アクティブ) ピクセルの数を  $\#b = \sum_{(x, y, t) \in I_b} 1$  とする。

20

【 0 0 4 2 】

我々は、セグメント  $B$  のサブセット選択および時間的なシフト  $M$  の費用を測定するエネルギー関数を定める。費用は、活動の損失  $E_a$ 、オブジェクト  $E_o$  間の重なりのパナルティ、および長い映像概要にパナルティを課す項  $E_l$  を含む。

$$E(M, B) = E_a + \alpha E_o + \beta E_l \quad (6)$$

30

ここで、

$$E_a = \sum_b \#b - \sum_{b \in B} \#b \quad (7)$$

$$E_o = \sum_{(x, y, t) \in S} Var\{src(x, y, t)\}$$

$$E_l = length(S)$$

40

である。

【 0 0 4 3 】

3. 1. 所定の長さを有する映像 - 概要

ここで、所定の長さ  $K$  の短い映像概要が、長い映像で構成される事例を説明する。このスキームの場合、各オブジェクトは、長さ  $K$  を有する重複し連続するセグメントに分割される。総てのセグメントは、時間的に推移してタイム  $t = 1$  で開始し、いずれのセグメントが映像概要内に含まれるか決定する必要がある。明らかに、このスキームでは、いくつかのオブジェクトが概要データ内に現れない。

【 0 0 4 4 】

初めに、我々はセグメントの総ての対の重なりを決定する。  $b_i$  および  $b_j$

50

を出現時間  $t_i$  および  $t_j$  を有する 2 つのセグメントとし、各セグメントのサポートを (等式 5 のような) 特有の関数により表わす。

【 0 0 4 5 】

これらの 2 つのセグメントの費用は、時間  $t = 1$  になった後の 2 つのセグメントの色差の合計として決定される。

$$v(b_i, b_j) = \sum_{x, y, t \in S} (I(x, y, t + t_i) - I(x, y, t + t_j))^2 \cdot \chi_{b_i}(x, t, t + t_i) \cdot \chi_{b_j}(x, t, t + t_j) \quad (8)$$

10

【 0 0 4 6 】

映像概要の場合、我々は、等式 6 の費用を最小化するセグメント B の部分的なセットを選択し、ここでは、 $E_1$  は、定数 K であり、重なり費用は、

$$E_o(B) = \sum_{i, j \in B} v(b_i, b_j) \quad (9)$$

により求められる。

20

【 0 0 4 7 】

同じ空間および時間ピクセルが 2 回表示される (これは、許容可能であるが無駄である) のを阻止すべく、我々は、オリジナルの動画を横切るセグメント  $b_i$  および  $b_j$  のために、 $v(b_i, b_j) =$  と設定する。さらに、ストロボ効果が必要ない場合、ストロボ効果は、同じオブジェクトから標本化された総ての  $b_i$  および  $b_j$  のために、 $v(b_i, b_j) =$  と設定することにより避けられる。

【 0 0 4 8 】

擬似的なアニーリング (simulated annealing) [ 8 ] は、エネルギー関数を最小化するために利用される。各状態は、概要内に含まれるセグメントのサブセットを示しており、隣接する状態が、設定するために取得され、セグメントが取り除かれ、追加され、あるいは別のセグメントと交換される。

30

【 0 0 4 9 】

セグメントの選択の後、長さ K の概要動画は、総ての変化したセグメントを結合することにより構成される。この手段を利用した映像概要の一つのフレームの例が、図 4 に示されている。

【 0 0 5 0 】

### 3.2. 無損失映像概要

映像監視などのいくつかの応用例の場合、我々は、長い映像概要を好むが、総ての活動が現れるのが保証される。この場合、オブジェクトは、先行するセクションで行われた一組のオブジェクトのセグメントを選択するためのものでないが、オブジェクトのセグメントのコンパクトな時間的な再配置を発見するためのものである。

40

【 0 0 5 1 】

さらに、我々は、擬似アニーリングを用いてエネルギーを最小化する。この場合、状態は、総てのセグメントの一組のタイムシフトに対応し、2 つの状態は、タイムシフトが単一のセグメントのためにのみ変化する場合、隣接部分として定義される。この場合、注意すべき 2 つの問題がある。

・最初あるいは最後のフレーム内で現れるオブジェクトのセグメントは、映像概要内に残るべきである (さもなければ、これらは突然現れるか消える)。我々は、各状態が、これら総てのオブジェクトの時間的なシフトを結果的に固定することにより、この制約を満足させることに留意する。

50

・入力映像の時間的な配置は、共通するエネルギー関数の局所的な最小化であり、したがって、アニメーション処理を初期化する最適な選択肢ではない。我々は、短い映像を有する疑似アニメーションを初期化し、総てのオブジェクトが重なる。

【 0 0 5 2 】

図 5 a , 5 b および 5 c は、短い概要が、ストロボ効果がなく活動の損失がない長いシーケンスを表示可能である場合に、この手段を示す図である。3つのオブジェクトは、同時に表示されるようにタイムシフトできる。特に、図 5 a は、オリジナルの映像（上図）および映像概要（下図）の空間および時間的な概略図である。図 5 b は、オリジナルの映像の3つのフレームを示しており、図 5 a に示すように、オリジナルの映像では、各人物が個別に現れているが、映像概要では、3つの総てのオブジェクトが一緒に現れる。図 5 c は、3人を同時に示す映像概要の1つのフレームを示している。

10

【 0 0 5 3 】

#### 4 . パノラマの映像概要

ビデオカメラがシーンをスキャンする場合、パノラマのモザイクを用いて多くの冗長な部分が取り除かれる。しかし、従来の方法は、シーンの動的性が損なわれる単一のパノラマ画像を構成する。制限された動的性は、ストロボ画像により表わすことができ [ 6 , 1 , 3 ]、移動するオブジェクトは、これらの経路に沿っていくつかの位置で表示される。

【 0 0 5 4 】

パノラマの映像概要は、異なる時間にシーンの異なる位置で行われる活動を同時に表示させることにより生成できる。実質的な圧縮が得られ、各オブジェクトの活動の期間が時間に制限されるため、カメラにより表示される。特別な事例は、カメラが、図 6 に示す走っているライオンなどのオブジェクトをトラックする場合である。カメラが、走っているライオンをトラックする場合、映像概要は、背景のパノラマのモザイクであり、前景は、走っているライオンのいくつかの動的なコピーを含む。この場合、短い映像概要は、ストロボ効果を許可するだけで取得できる。

20

【 0 0 5 5 】

パノラマの映像概要を構成することは、総てのフレームをいくつかの基準フレームに整列させる予備の段階で、通常の映像概要と同様の方法により行われる。整列の後、オブジェクトの画像の座標が、入力画像のうちの一つの座標系であるグローバル座標から取得される。

30

【 0 0 5 6 】

映像を処理を可能にすべく、移動するオブジェクトのセグメンテーションが完全でない場合でさえ、我々は、重なった部分を完全に阻止する代わりに当該部分にペナルティを課す。この重なった部分のペナルティにより、オブジェクトの時間的な配置の順応性を可能にし、セグメンテーションが完全でない場合でさえ、オブジェクトのピクセルは、いくつかの背景を含む。

【 0 0 5 7 】

付加的な期間が追加され、映像概要の時間的な順序を入力映像の順序に偏向させる。

【 0 0 5 8 】

総ての可能なセグメントセクション B および時間的なシフト M の前述したエネルギーを最小化することは、可能性が非常に多いため、非常に消耗的である。しかしながら、問題は、解決手段を制限することにより小さくすることができる。2つの制限されたスキームは、以下のセクションに記載される。

40

【 0 0 5 9 】

#### 5 . 監視例

映像概要の関心のある応用例は、記録された監視映像へのアクセスである。映像内の特定のイベントを調査する必要がある場合、調査は、映像概要を用いることにより非常に早く行うことができる。

【 0 0 6 0 】

前述したように、図 5 は、いずれの活動を失うことなく総ての活動を短い期間に凝縮し

50

た映像概要のパワーの例を示している。これは、コーヒーステーションを監視するカメラが収集した映像を用いて行われた。2つの例は、実際の監視カメラから得られたものである。図8 a, 8 bおよび8 cは、ストリートの監視の映像概要の詳細を示す図である。図8 aは、オリジナル映像の通常のフレーム(22秒)を示している。図8 bは、凝縮された活動を示す映像概要動画(2秒)のフレームを示している。図8 cは、短い映像概要(0.7秒)のフレームを示しており、さらに凝縮された活動を示している。これらの図に示す画像は、街路を監視するカメラにより取得された映像から生成され、歩行者が時折視界を横切っている。これらの多くは、非常に凝縮された概要に集められる。

#### 【0061】

図8 aおよび8 bは、フェンスの監視の映像概要の詳細を示す図である。フェンスの近くでは非常に少ない活動があり、時折、兵士が巡回するのが確認できる。映像概要は、巡回および歩いている兵士の総ての事例を同時に示しており、任意で、ストロボ効果により表示させることにより映像概要を短くすることさえ可能である。

#### 【0062】

### 6. 映像概要による映像インデックス

映像概要は、映像インデックスとして利用することができ、映像内の行動にアクセスする効率的で直観的なリンクをユーザに提供する。これは、総ての概要ピクセルを、オリジナル映像の対応するオブジェクトの外観へのポイントと関連させることにより行うことができる。映像概要では、映像の情報は、「活動の空間」に投影され、活動は、時間的な内容に拘わらず(我々は空間的な内容を保護できるが)重要である。活動は短い期間内に集中するため、映像内の特定の活動は容易にアクセスできる。

#### 【0063】

前述した説明から、ビデオカメラが動的なシーンをスキャンニングする場合、領域が入力映像内で確認できるようになる絶対的な「年代順の時間」が、シーンの動的性の一部でないことは理解できるであろう。各領域の可視期間内の「ローカルタイム」は、シーン内の動的性の描写とより関連しており、動的なモザイクを構成する場合に保護されるべきである。前述した実施例は、本発明の第1の態様である。第2の態様では、我々は、継ぎ目のないパノラマのモザイクを生成する方法を示しており、当該実施例では、シーン内のオブジェクトが移動する場合に、画像間のステッチが、可能な限りシーン内のオブジェクトの部分省略することを阻止する。

#### 【0064】

### 7. 3次元の最小カットを用いたパノラマ画像の生成

$I_1, \dots, I_N$  は、入力シーケンスのフレームである。我々は、シーケンスが、従来の方法のうちの一つを用いて、単一の基準フレームに整列された場合を想定する。説明を簡単にするために、我々は、整列後の総てのフレームが、同じ大きさであると仮定する(カメラの視界の外のピクセルは、無効として記録される)。また、カメラは時計回りにパンすると仮定する(別の動作は、同じような方法で処理できる)。

#### 【0065】

$P(x, y)$  は、構成されたパノラマ画像とする。 $P$ 内の各ピクセル $(x, y)$ の場合、我々は、このピクセルが取得されるフレーム $M(x, y)$ を選択する必要がある(すなわち、 $M(x, y) = k$ の場合、 $P(x, y) = I_k(x, y)$ である)。カメラが時計回りにパンすると仮定した場合、左側の列は最初のフレームから取得され、一方、右側の列は、最後のフレームから取得されることは明らかである(視野の小さなパノラマ画像を生成するために、他の境界条件を選択できる)。

#### 【0066】

我々の目的は、継ぎ目のないパノラマ画像を生成することである。このため、我々は、特に、オブジェクトが移動している場合、オブジェクト内のステッチの防止を試みる。我々は、[1]で使用されるスコアと同様の継ぎ目のスコアを利用するが、NPの難しい問題を(近似により)解決する代わりに、より限定された問題の最適な解決方法を発見する。

10

20

30

40

50

【 0 0 6 7 】

8 . エネルギー最小化問題などの問題の定式化

前述した等式との主な違いは、

$$E_{stitch}(x, y, x', y') = \sum_{k=\min M}^{\max M-1} \frac{1}{2} \|I_k(x, y) - I_{k+1}(x, y)\|^2 + \frac{1}{2} \|I_k(x', y') - I_{k+1}(x', y')\|^2 \quad (10)$$

により求められるステッチ費用である。ここで、

min = min ( M ( x , y ) , M ( x ' , y ' ) )、

max = max ( M ( x , y ) , M ( x ' , y ' ) ) である。

【 0 0 6 8 】

この費用は、フレームの割り当てが連続的であると仮定すると妥当であり、すなわち、( x , y ) および ( x ' , y ' ) が隣接するピクセルの場合、これらのソースフレーム M ( x , y ) および M ( x ' , y ' ) は近い。この費用の主な利点は、この問題を図上の最小カットの問題として解決できることである。

【 0 0 6 9 】

我々が最小化するエネルギー関数は、

$$E(M) = \sum_{(x,y)} \sum_{(x',y') \in N(x,y)} E_{stitch}(x, y, x', y') + \sum_{(x,y)} (1 - Valid(x, y, M(x, y))) \cdot D, \quad (11)$$

であり、ここでは、

N ( x , y ) は、( x , y ) の隣接するピクセルである。

E ( x , y , x ' , y ' ) は、等式 1 で説明したように、隣接する各ピクセルのステッチ費用である。

Valid ( x , y , k ) は 1 であり、I<sub>k</sub> ( x , y ) は、有効ピクセル ( 即ち、カメラの視界内 ) である。

D は、非常に大きな数である ( 無限大を意味する ) 。

【 0 0 7 0 】

9 . 単一のパノラマの生成

次に、我々は、( 複雑な指数を有する ) 2 次元の複数のラベルの問題を、( 複雑な多項式を有するが実際には即座に解決できる ) 3 次元の 2 つのラベルの問題に変換する方法を示す。各ピクセル x , y および入力フレーム k の場合、我々は、X ( x , y ) ≤ k の場合に 1 に等しい 2 値変数 b ( x , y , k ) を決定する ( M ( x , y ) は、ピクセル ( x , y ) のソースフレームである ) 。 b ( x , y , k ) = 1 であることは明らかである。

【 0 0 7 1 】

各 1 ≤ k ≤ N で b ( x , y , k ) の場合、我々は、b ( x , y , k ) = 1 の最小の k として M ( x , y ) を定める。我々は、最小化が継ぎ目のないパノラマを与えるエネルギー項を記載する。各隣接するピクセル ( x , y ) および ( x ' , y ' ) と、各 k の場合、我々は、b ( x , y , k ) b ( x ' , y ' , k ) の割り当てのために誤差項 ( error term ) を追加する ( この誤差項は対称である ) 。

$$\|I_k(x, y) - I_{k+1}(x, y)\|^2 + \|I_k(x', y') - I_{k+1}(x', y')\|^2$$

10

20

30

40

50

## 【 0 0 7 2 】

また、我々は、 $b(x, y, k) = 1$ であるが $b(x, y, k + 1) = 0$ である割り当ての無限のペナルティを追加する( $M(x, y) \leq k$ であるが $M(x, y) > k$ が可能でないため)。

## 【 0 0 7 3 】

最後に、 $I_k(x, y)$ が無効なピクセルである場合、我々は、無限のペナルティを割り当てに与えることにより、 $k > 1$ の場合 $b(x, y, k) = 1$   $b(x, y, k + 1) = 0$ あるいは $k = 1$ の場合 $b(x, y, k) = 1$ 、このピクセルを選択することを阻止することができる(これらの割り当ては、 $M(x, y) = k$ であることを意味している)。

## 【 0 0 7 4 】

前述した総ての項は、3次元グリッドの変数の対であり、したがって、我々は、3次元バイナリMRFのエネルギー関数を最小化として説明でき、最小カットを用いて多項式時間内でそれを最小化できる[9]。

## 【 0 0 7 5 】

10 . 4次元最小カットを用いたパノラマ動画の生成

(長さLの)パノラマ動画を生成すべく、我々は、一連のパノラマ画像を生成すべきである。時間的な整合性が強化されないため、各パノラマ画像を個別に生成することはよくない。別の方法は、連続的なモザイク画像が、先行するモザイクで使用された連続するフレームから各ピクセルを取得する場合、最初のフレームとして最初のモザイク画像で開始する( $M_1(x, y) = M(x, y) + 1$ )。この可能性は、図2bを参照して前述したのと同様である。

## 【 0 0 7 6 】

本発明の第2の態様では、我々は、1のパノラマフレームから別のパノラマフレームに変化する機会をステッチに与える別の式を代わりに使用し、これは、移動するオブジェクトを上手くステッチするのに非常に重要である。

## 【 0 0 7 7 】

我々は、前述した3次元図の事例Lで構成される4次元の図を生成する。

$$b(x, y, k, l) = 1 \iff M_l(x, y) \leq k.$$

## 【 0 0 7 8 】

時間的な整合性を強化すべく、我々は、無限のペナルティを割り当てに課す、即ち $l < L$ それぞれの場合は $b(x, y, N, l) = 1$ 、 $l > 1$ それぞれの場合は $b(x, y, 1, l) = 0$ 。

## 【 0 0 7 9 】

さらに、各 $(x, y, k, l)$  ( $1 \leq l \leq L - 1, 1 \leq k \leq N - 1$ )の場合、我々は、割り当て $b(x, y, k, l) = 1$   $b(x, y, k + 1, l + 1)$ のための費用関数を設定する( $k = N - 1$ の場合、我々は費用の左項のみを使用する)。

$$E_{temp} = \frac{1}{2} \|I_k(x, y) - I_{k+1}(x, y)\|^2 + \frac{1}{2} \|I_{k+1}(x, y) - I_{k+2}(x, y)\|^2 \quad (12)$$

この費用は、生成された動画内で(時間的に)連続したピクセルを表示するのを強化する(ただし、例えば、これらのピクセルは背景内である)。

## 【 0 0 8 0 】

この方法の変更例は、各ピクセル $(x, y)$ を、連続したフレームの同じピクセルに接

10

20

30

40

50

続するのではなく、ピクセル (  $u, v$  ) におけるオプティカルフローにより、対応するピクセル (  $x + u, y + v$  ) に接続する。オプティカルフローを算出する好適な方法は、例えば [ 19 ] にある。オプティカルフローを用いることにより、移動するオブジェクトの事例をよりよく処理する。

#### 【 0081 】

さらに我々は、4次元図の最小カットを用いてエネルギー関数を最小化することができ、バイナリの解決方法は、ステッチの問題を軽減するパノラマ動画を決定する。

#### 【 0082 】

##### 11. 実用的な改良

4次元の図を保存するには、非常に大きなメモリが必要である。したがって、我々は、メモリ要求とアルゴリズムの動作時間の双方を低減するいくつかの改良点を利用する。

- ・前述したように、エネルギーは、無効なピクセルのための頂点 ( vertices ) を明確に保存することなく最小化できる。したがって、頂点の数は、入力映像内のピクセルの数に減少し、出力映像内のフレームの数により増加する。

- ・出力映像内の各フレームを解決する代わりに、我々は、出力フレームの標本化されたセットのみを解決し、これらの間にステッチ関数を挿入する。この改良点は、シーン内の動作があまり大きくないことを想定している。

- ・我々は、入力フレームの部分的なセットからのみ各ピクセルを生成することができる。これは、特に映像から取得される一連のフレームに有用であり、連続したフレームの各対の間の動作は非常に小さい。この場合、我々は、各ピクセルのために一組のソースフレームを標本化することにより、多くを失うことがない。しかし、一貫性のある方法によりソースフレームを標本化することが好適である。例えば、フレーム  $k$  が、 $1 - t$  h 出力フレーム内のピクセル (  $x, y$  ) のために可能なソースである場合、 $k + 1$  フレームは、 $1 + 1 - t$  h 出力フレーム内のピクセル (  $x, y$  ) のために可能なソースフレームであるべきである。

- ・我々は、複数の解像度のフレームワーク ( multi-resolution framework ) を利用し ( 例えば、[ 2 ] で利用されたような )、粗い解決 ( solution ) が低解像度の画像のために発見され ( バッファリングおよび標本化の後に )、解決は境界においてのみ洗練される。

#### 【 0083 】

ここで、我々は、関心のあるスコアにより動画を結合する方法を説明する。密度の高い ( denser ) あるいは薄い ( sparser ) 活動を含む動画を生成し、あるいはユーザが特定する方法でシーンを制御するといったいくつかの応用例が存在する。

#### 【 0084 】

[ 14 ] に記載された動的なパノラマは、特別な事例として考えることができ、同じ動画の異なる部分が結合されて、より大きな視界の動画が取得され、この場合、我々は、各時間の各ピクセルの「視認性」により、関心のあるスコアを決定する。さらに一般的には、同じ動画の異なる部分 ( 時間あるいは空間的にシフトする ) を結合することは、他の事例でも利用できる。例えば、動画内の活動の密度を高めるために、我々は、アクションが生じる動画の異なる部分を、多くのアクションを有する新しい動画に結合する。図 1 から 8 を参照して前述した実施例は、活動の特別な事例を説明しており、異なる方法を利用している。

#### 【 0085 】

取り組むべき2つの問題は、

1. 動画を「見た目のよい」動画に結合する方法。例えば、我々はステッチの問題を避けたい。

2. 関心のあるスコアを最大化する。

#### 【 0086 】

我々は、使用可能な異なるスコアを説明することから開始し、動画を結合するのに用いられるスキームを説明する。

#### 【 0087 】

10

20

30

40

50

関心のある特徴 (interest function) として使用可能な主要な特徴の一つは、ピクセルの「重要度」のレベルである。我々の実験では、我々は、ピクセル内の「活動」をその重要度を示すために考慮しているが、重要度の他の測定も同様に好適である。活動レベルの評価は、本発明の特徴自体ではないが、前述したセクション 1 で参照された様々な方法のうちの一つ (活動検出) を利用して行うことができる。

【 0 0 8 8 】

### 1 3 . 他のスコア

動画を結合するのに利用可能な他のスコア :

- ・視認性スコア : カメラが移動する場合、あるいは映像内の穴を埋めようとする場合、見ることができないピクセルが存在する。我々は、(必ずしも無限のスコアによってではなく) 無効なピクセルにペナルティを課することができる。この方法では、我々は、穴を埋めることを促進できるが (あるいは視野を広げることができるが)、悪いステッチが生じた場合には、穴を埋めることを望まなくてもよく、あるいは小さな視野を利用してよい。

10

- ・方向 : 活動の測定は、方向の測定と置き換えることができる。例えば、我々は、垂直に移動する領域よりも水平に移動する領域を好むだろう。

- ・ユーザ仕様 : ユーザは、色、テクスチャ等の好適な関心のある特徴を指定する。さらに、ユーザは、異なるスコアにより手動で領域 (およびタイムスロット) を指定する。例えば、1 が、最大の活動が所望されることを意味し、0 が、活動がないことが所望されることを意味するマスクを描くことにより、ユーザは、シーン内の動的性を制御でき、すなわち、特定の位置で発生させることができる。

20

【 0 0 8 9 】

### 1 4 . アルゴリズム

我々は、以下の変更を有する [ 2 0 ] により使用される方法と同様の方法を利用する。

- ・我々は、1 の動画あるいは別の動画から選択される各ピクセルのための関心のある領域を追加する。このスコアは、各動画の各ピクセルのエッジを利用して端末の頂点 (ソースおよびシンク) に追加され、これらのエッジの重みは、関心のあるスコアである。

- ・我々は、(任意で) フレームの連続した各対の間のオプティカルフローを算出する。次に、一貫性を強化すべく、我々は、時間的に隣接する部分のエッジ ( ( x , y , t ) から ( x , y , t + 1 ) ) を、オプティカルフローによる隣接部分のエッジ ( ( x , y , t ) から ( x + u ( x , y ) , y + v ( x , y ) , t + 1 ) ) と交換できる。これは、ステッチが顕著でないフローに従うように奨励するように、ステッチ動画間の遷移を強化する。

30

- ・動画のいずれの部分 (あるいは移動する部分) が結合されるかを決定する場合、ステッチ費用だけでなく、関心のあるスコアを考慮すべきである。例えば、密度の高い活動レベルを有する動画を生成する場合、我々は、スコアを最大化する動画のセット S を選択する。

$$\sum_{x,y,t} \bigcup_{b \in S} \chi_b(x,y,t)$$

40

【 0 0 9 0 】

図 9 b は、動画の増加した活動密度のような効果を示した図であり、オリジナルのフレームは図 9 a に示されている。2 つ以上の動画が結合された場合、我々は反復手段を利用し、この反復により、新しい動画が、生成された動画に結合される。正確に行うために、以前の反復により生成された古い継ぎ目およびスコアを考慮すべきである。関心のあるスコアが無いにも拘わらず、このスキームは [ 2 0 ] に記載されている。生成された映像のサンプルフレームが、図 9 b に示されている。

【 0 0 9 1 】

50

図10は、処理の概略図である。この例では、映像は、時間的にシフトした映像と結合される。結合は、前述した基準、すなわちステッチ費用を最小限にしつつ関心のあるスコアを最大化することに基づいて、最小カットを用いて行われる。

#### 【0092】

ここで図11を参照すると、カメラ11により取得された第1の動的シーンの映像フレームの第1のシーケンスを、第2の動的シーンを示す少なくとも2つの映像フレームの第2のシーケンスに変換する本発明に係るシステム10のブロック図が示されている。このシステムは、各x, y座標に配置された複数のピクセルを備える少なくとも一つのオブジェクトの動作を示す第1のシーケンス無いの映像フレームのサブセットを記憶する第1のメモリ12を備えている。第1の動的なシーン内の少なくとも一つのオブジェクトの空間的に重ならない外観を示すサブセット部分から選択する選択ユニット13が、第1のメモリ12に連結される。フレーム生成部14は、オブジェクト内のピクセルの各x, y座標を変更することなく、少なくとも3つの異なる入力フレームの部分を、第2のシーケンスの少なくとも2つの連続したフレームにコピーし、第2のシーケンスの少なくとも一つのフレームが、第1のシーケンス内の異なるフレームに現れる少なくとも2つの部分を含む。第2のシーケンスのフレームは、ディスプレイユニット16による後の処理あるいは表示のために、第2のメモリ15に格納される。フレーム生成部14は、第2のシーケンスをコピーする前に少なくとも2つの部分を空間的に移動させる移動ユニット(warping unit)17を備えてもよい。

10

#### 【0093】

実際には、システム10は、本分野で周知のように、グラフィックカードあるいはワークステーションおよび好適な周辺機器を有する最適にプログラムされたコンピュータにより実現される。

20

#### 【0094】

システム10では、少なくとも3つの異なる入力フレームは時間的に近い。システム10はさらに、映像フレームの第1のシーケンスを予め整列させるために、第1のメモリに結合された任意の整列ユニット18を備えてもよい。この場合、カメラ11は、第1のメモリ12内の予め整列された映像フレームを格納するために整列ユニット18に連結される。整列ユニット18は、

第1のシーケンス内のフレーム間の画像動作パラメータを算出し、

第1の動的シーン内の静止したオブジェクトが映像内で静止するように、第1のシーケンス内の映像フレームを移動させることにより動作する。

30

#### 【0095】

同様に、システム10は、整列された空間および時間量を「タイムフロント」表面により一掃し、一連のタイムスライスを生成するために、選択ユニット13に連結された任意のタイムスライス生成部19を備えてもよい。

#### 【0096】

これらの任意の特徴は、「タイムフロント」および「タイムスライス」の語は、参照されている前述したWO2006/048875に完全に記載されているのと同様であるため、詳細には説明されていない。

40

#### 【0097】

完全にするために、図12は、本発明に係るシステム10により実行される主要な処理を示すフロー図である。

#### 【0098】

##### 15. 考察

映像概要は、映像内の活動を非常に短い期間に凝縮する手段として提案されている。この凝縮された表示は、映像シーケンス内の活動への効率的なアクセスを可能にする。2つの手段が提案されており、1つの手段は、低いレベルのグラフの最適化を用い、映像概要内の各ピクセルは、このグラフ内のノードである。この手段は、入力映像から直接的に映像概要を取得する利点を有しているが、解決(solution)の複雑度が非常に高い。代替的

50

な手段は、最初に移動するオブジェクトを検出し、検出されたオブジェクトを最適化することである。第2の手段では動作のセグメンテーションの予備段階が必要とされるが、これは非常に高速であり、オブジェクトに基づく制約が可能である。生成される映像概要の活動は、オリジナルの映像よりも非常に凝縮されており、このような概要は、これを利用したことの無い者にとっては扱い難いかもしれない。しかし、多くの情報を短い時間内で観察することを目的とする場合、映像概要はこの目的を達成する。特別な注意が、動的なストロボスコープを取得する可能性に払われるべきである。映像概要の長さをさらに減少させるとともに、動的なストロボスコープはユーザからの調整を必要としてもよい。単一のオブジェクトの複数の空間的な出現は、長い活動時間を示すことが分かることにはいくらかの訓練を必要とする。我々は、動的な映像概要の特定の実施例を説明しているが、多くの拡張も可能である。例えば、バイナリの「活動」インジケータを有するのではなく、活動インジケータが連続していてもよい。連続的な活動は、例えば、活動レベルに基づいて、表示されたオブジェクトの速度を制御することにより、映像概要を生成可能な選択肢を広げることができる。また、映像概要は、多くのショットから成る長い動画のために用いてもよい。理論的に、我々のアルゴリズムは、重なり（あるいは不連続性）による不利益のために、異なるシーンの部分を結合させない。この場合、単一のショットに利用される単一の背景モデルは、調整可能な背景推定器と交換される。長い動画に利用される別の手段は、ショットのバイナリ検出の従来の方法を利用し、ショット毎に個別に映像概要を生成することである。

10

【0099】

20

本発明に係るシステムは、プログラムされたコンピュータに最適であることは理解できるであろう。同様に本発明は、本発明の方法を実行すべく、コンピュータにより読み取り可能なコンピュータプログラムを意図する。さらに本発明は、本発明の方法を実施する機械により実行可能なプログラム命令を明確に実行する機械読み取り可能なメモリを意図する。

【図面の簡単な説明】

【0100】

【図1】図1は、時間的に配置された特徴を同時に表示することによりコンパクトな映像概要を生成する本発明のアプローチを示す図である。

【図2a】図2aは、本発明により生成される映像概要を表わす概略図である。

30

【図2b】図2bは、本発明により生成される映像概要を表わす概略図である。

【図3a】図3aは、本発明による一時的な再配置の例を示す図である。

【図3b】図3bは、本発明による一時的な再配置の例を示す図である。

【図3c】図3cは、本発明による一時的な再配置の例を示す図である。

【図4】図4は、図3bに示す動的なストロボ効果を利用した映像概要の単一のフレームを示す図である。

【図5a】図5aは、短い概要が、活動を損なうことなく且つストロボ効果なしで、長いシーケンスを表示できる場合の例を示す図である。

【図5b】図5bは、短い概要が、活動を損なうことなく且つストロボ効果なしで、長いシーケンスを表示できる場合の例を示す図である。

40

【図5c】図5cは、短い概要が、活動を損なうことなく且つストロボ効果なしで、長いシーケンスを表示できる場合の例を示す図である。

【図6】図6は、本発明によるパノラマの映像概要の更なる例を示す図である。

【図7a】図7aは、ストリートの監視の映像概要の詳細を示す図である。

【図7b】図7bは、ストリートの監視の映像概要の詳細を示す図である。

【図7c】図7cは、ストリートの監視の映像概要の詳細を示す図である。

【図8a】図8aは、フェンスの監視の映像概要の詳細を示す図である。

【図8b】図8bは、フェンスの監視の映像概要の詳細を示す図である。

【図9a】図9aは、本発明の更なる実施例による動画の増加する活動密度を示す図である。

50

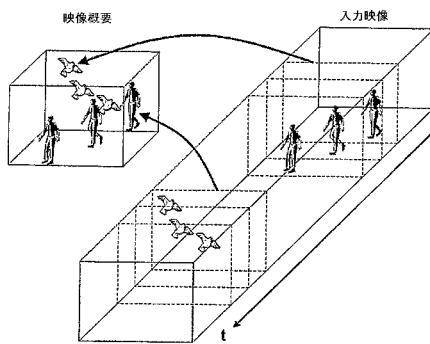
【図9b】図9bは、本発明の更なる実施例による動画の増加する活動密度を示す図である。

【図10】図10は、図9に示す動画を生成するのに利用される処理の概略図である。

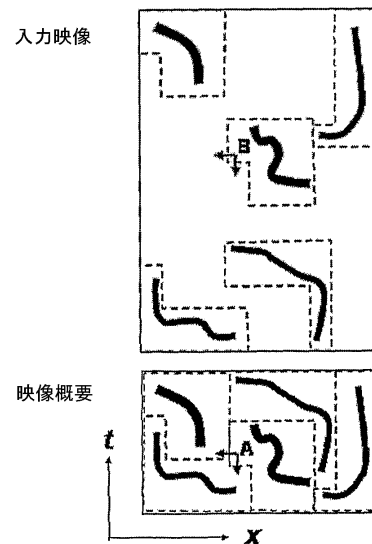
【図11】図11は、本発明によるシステムの主要な機能を示すブロック図である。

【図12】図12は、本発明により実行される主要な動作を示すフロー図である。

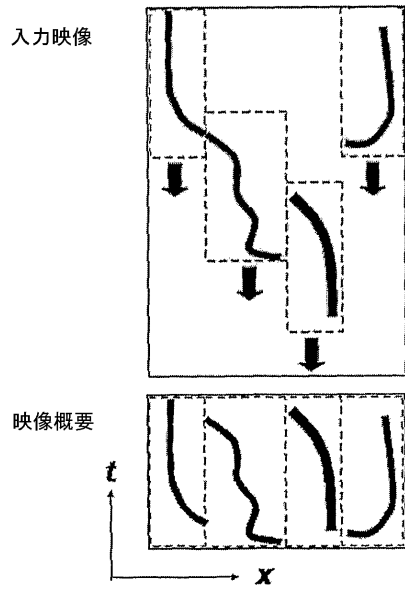
【図1】



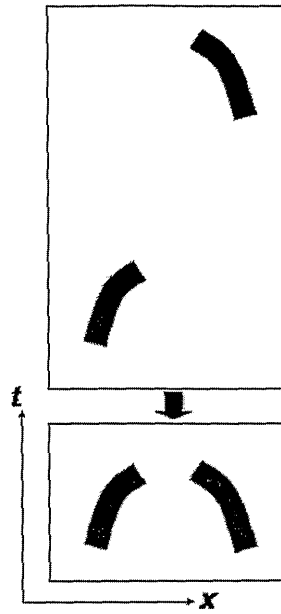
【図2a】



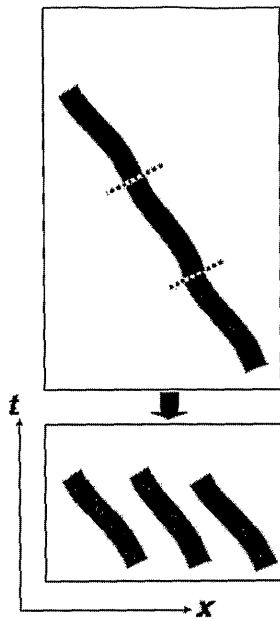
【図 2 b】



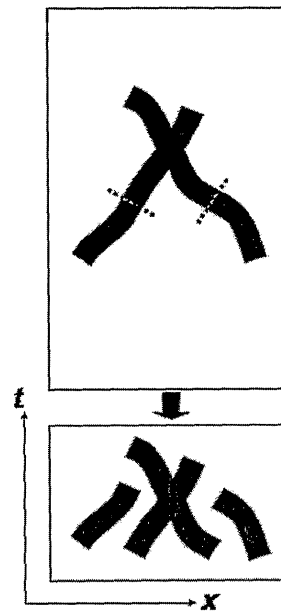
【図 3 a】



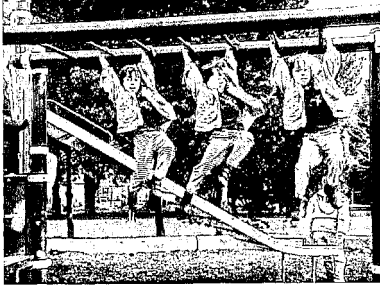
【図 3 b】



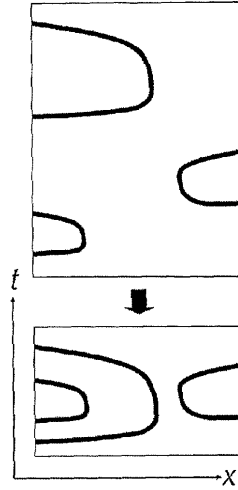
【図 3 c】



【 図 4 】



【 図 5 a 】



【 図 5 b 】



【 図 5 c 】



【 図 7 b 】



【 図 6 】



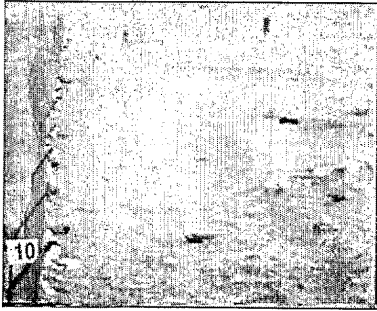
【 図 7 a 】



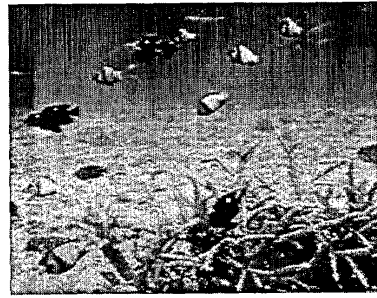
【 図 7 c 】



【図 8 a】



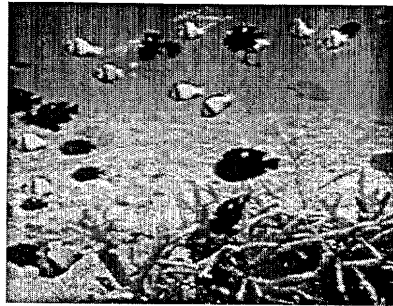
【図 9 a】



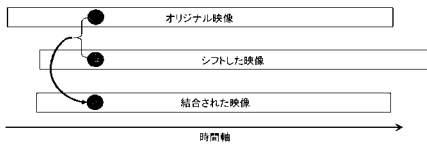
【図 8 b】



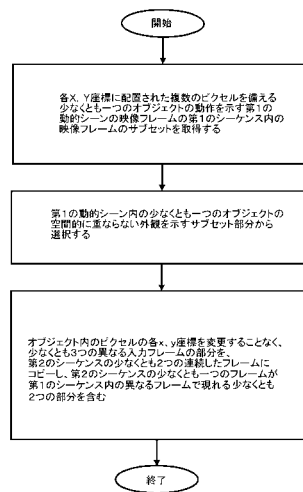
【図 9 b】



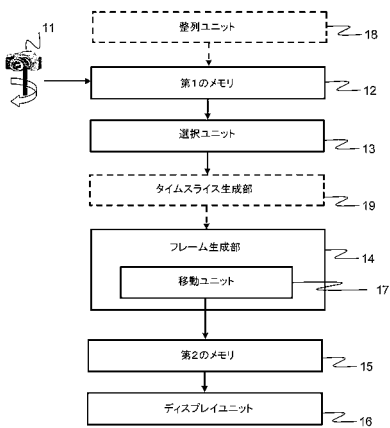
【図 10】



【図 12】



【図 11】



10

---

フロントページの続き

- (72)発明者 ペレグ, シュムエル  
イスラエル国 メヴァセレットジオン 90805, メロンストリート34/8
- (72)発明者 ラヴ - アチャ, アレクサンダー  
イスラエル国 エルサレム 96757, クボヴィストリート58

審査官 板垣 有紀

- (56)参考文献 特開2005 - 210573 (JP, A)  
国際公開第2004/040480 (WO, A1)  
国際公開第01/078050 (WO, A1)

(58)調査した分野(Int.Cl., DB名)

G06T 3/00  
G06T 7/20  
H04N 5/91  
H04N 5/46  
H04N 7/18