

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4446124号
(P4446124)

(45) 発行日 平成22年4月7日(2010.4.7)

(24) 登録日 平成22年1月29日(2010.1.29)

(51) Int.Cl.	F I		
HO4N 5/91 (2006.01)	HO4N	5/91	N
G06T 13/00 (2006.01)	HO4N	5/91	R
	G06T	13/00	B

請求項の数 2 (全 25 頁)

(21) 出願番号	特願2005-337185 (P2005-337185)	(73) 特許権者	393031586
(22) 出願日	平成17年11月22日(2005.11.22)		株式会社国際電気通信基礎技術研究所
(65) 公開番号	特開2007-143029 (P2007-143029A)		京都府相楽郡精華町光台二丁目2番地2
(43) 公開日	平成19年6月7日(2007.6.7)	(74) 代理人	100090181
審査請求日	平成20年3月31日(2008.3.31)		弁理士 山田 義人
特許法第30条第1項適用	2005年5月24日 社	(72) 発明者	桑原 教彰
団法人電子情報通信学会発行の「電子情報通信学会技術			京都府相楽郡精華町光台二丁目2番地2
研究報告 信学技報Vol. 105 No. 105」に		(72) 発明者	株式会社国際電気通信基礎技術研究所内
発表			桑原 和宏
(出願人による申告)平成17年度独立行政法人情報通		(72) 発明者	京都府相楽郡精華町光台二丁目2番地2
信研究機構、研究テーマ「軽度脳障害者のための情報セ			株式会社国際電気通信基礎技術研究所内
ラピーインタフェースの研究開発」に関する委託研究、			安部 伸治
産業活力再生特別措置法第30条の適用を受ける特許出			京都府相楽郡精華町光台二丁目2番地2
願			株式会社国際電気通信基礎技術研究所内
			最終頁に続く

(54) 【発明の名称】 ビデオコンテンツ作成方法

(57) 【特許請求の範囲】

【請求項1】

写真遷移にフェードアウト/フェードインの視覚効果を用い、写真の再生期間中にナレーションおよび反応時間を配置したビデオコンテンツを作成する方法であって、

最小フェードアウト/フェードイン時間、ナレーション時間および反応時間の総和より最大ビデオ再生時間が小さいとき写真を除外し、そして

最大フェードアウト/フェードイン時間、ナレーション時間および反応時間の総和より最小ビデオ再生時間が大きいとき、写真を追加するようにした、ビデオコンテンツ作成方法

。【請求項2】

写真遷移にフェードアウト/フェードインの視覚効果を用い、同じ写真内での人から人へ遷移にパンズームの視覚効果を用い、写真の再生期間中にナレーションおよび反応時間を配置したビデオコンテンツを作成する方法であって、

最小フェードアウト/フェードイン時間、ナレーション時間、反応時間および最小パンズーム時間の総和より最大ビデオ再生時間が小さいとき写真を除外し、そして

最大フェードアウト/フェードイン時間、ナレーション時間、反応時間および最大パンズーム時間の総和より最小ビデオ再生時間が大きいとき、写真を追加するようにした、ビデオコンテンツ作成方法。

【発明の詳細な説明】

【技術分野】

【 0 0 0 1 】

この発明はビデオコンテンツ作成方法に関し、特にたとえば、認知症者の過去の写真をナレーションとともにその認知症社（視聴者）に提示できる、思い出ビデオのようなビデオコンテンツを作成する、ビデオコンテンツ作成方法に関する。

【背景技術】

【 0 0 0 2 】

高齢の認知症者に視聴覚刺激を提示することで、彼（女）らを活性化することを目的とした研究が幾つかなされている。たとえば、非特許文献 1 に示すビデオレスパイトでは、ビデオの視聴者である認知症者に語り掛けを行うキャラクタが提示される。また、非特許文献 2 に示すプロジェクト CIRCA では、昔の有名な歌や画像、映像を含んだ、良くデザインされたマルチメディアコンテンツが利用されている。我々は、認知症者の長期記憶を刺激するという観点から、視聴覚刺激の素材として、思い出ビデオを選定した。思い出ビデオは認知症者の古いアルバム中の写真を用いて作成されたスライドショービデオである。その臨床的な有効性は非特許文献 3 で実験的に示されている。

10

【 0 0 0 3 】

しかし、思い出ビデオを作成するのは単純な作業ではない。まず、古いアルバムを用いし、そこから適当な写真を選択する。使用する写真は、視聴者である認知症者の遠い昔の記憶を呼び覚ますものでなくてはならない。次に、ビデオカメラで写真をムービー仕立てに撮影するが、必要に応じてパン、ズームの映像効果（非特許文献 4 に示すいわゆるケンパーズ効果）を付与する。またナレーションは、認知症者をビデオにより一層引き付けるために付与される。

20

【 0 0 0 4 】

通常これらは、映像編集のスキルを有するボランティアが実施しており、介護家族が気軽に作成できるようなものではない。加えて、認知症者が思い出ビデオを繰り返し視聴することでそれに飽きてしまい、結果として思い出ビデオが当初のように認知症者を引き付ける効果は、次第に失せてしまうことも考えられる。

【非特許文献 1】Lund, D.A., Hill, R.D., Caserta, M.S., and Wright, S.D.: Video Respite: an innovative resource for family, professional caregivers, and persons with dementia, *The Gerontologist*, Vol. 35, Issue 5 (1995) 683-687.

【非特許文献 2】Gowans, G., Campbell, J., Alm, N., Dye, R., Astell, A., and Ellis, M.: Designing a multimedia conversation aid for reminiscence therapy in dementia care environments, *Extended abstracts of the 2004 conference on Human Factors and Computing Systems* (2004) 825 - 836.

30

【非特許文献 3】安田ほか：認知症者への思い出写真ビデオの作成と集中度の評価。第 28 回高次脳機能障害学会総会（2004）

【非特許文献 4】http://en.wikipedia.org/wiki/Ken_Burns

【発明の開示】

【発明が解決しようとする課題】

【 0 0 0 5 】

魅力的な思い出ビデオを作成するために、どのような映像効果、あるいはオーディオ効果が加えられるべきかを明確にする必要がある。そのため発明者等は、3つの代表的な効果、すなわち前述のケンパーズ効果、BGM、およびナレーションに対して、それらの有効性を評価するための、以下のような実験を実施した。上記のすべての効果を使用した思い出ビデオと、それら効果の1つを除外したビデオを、3人の認知症者のために用意した。それぞれのビデオの長さはおよそ20分である。2週間の実験期間中に介護家族には、毎日1回、映像効果、オーディオの効果を変え、上記すべてのタイプのビデオを認知症者が視聴するように依頼した。認知症者がビデオに飽きた場合には、ビデオを別のタイプに変えるように依頼した。介護家族には、認知症者がどの程度の長い間、そのタイプのビデオを視聴したか、またどの程度集中してそれを視聴したか、「1」から「5」の尺度で介護家族の主観で評価し記録してもらった。実験期間が終わった後に、介護家族のニーズ

40

50

を調査するために、インタビューも実施した。
その結果、以下のような知見が得られた。

【 0 0 0 6 】

(1) ナレーションは、魅力的な思い出ビデオに非常に有効である。

【 0 0 0 7 】

(2) 写真中のよく知った人々の顔に対して、ズームアップすることも不可欠である。

【 0 0 0 8 】

(3) さらに、ナレーションとズームが連動していることが重要である。すなわち、ナレーションが写真中の特定の人に関するものなら、ナレーションはその人のズームアップの際に付与されなければならない。

10

【 0 0 0 9 】

このように、思い出ビデオは認知症者の安定した精神状態を形成する上で有効性が確認されているが、特に、映像効果を付与したりナレーションを付与したりすることは重要な要素であり、さらにそれらは上述のように同期させる必要がある。

【 0 0 1 0 】

それゆえに、この発明の主たる目的は、新規な、ビデオコンテンツ作成方法を提供することである。

【 0 0 1 1 】

この発明の他の目的は、映像効果要素やナレーション要素などの各要素が同期したビデオコンテンツを作成できる、ビデオコンテンツ作成方法を提供することである。

20

【課題を解決するための手段】

【 0 0 1 2 】

請求項1の発明は、写真遷移にフェードアウト/フェードインの視覚効果を用い、写真の再生期間中にナレーションおよび反応時間を配置したビデオコンテンツを作成する方法であって、最小フェードアウト/フェードイン時間、ナレーション時間および反応時間の総和より最大ビデオ再生時間が小さいとき写真を除外し、そして最大フェードアウト/フェードイン時間、ナレーション時間および反応時間の総和より最小ビデオ再生時間が大きいとき、写真を追加するようにした、ビデオコンテンツ作成方法である。

【 0 0 1 3 】

請求項1の発明では、たとえばイメージスキャナなどを含む写真データ入力装置(20。実施例で相当する部分または要素を例示する参照符号。以下同様。)から、写真画像データをコンピュータ(12)に入力する。コンピュータはその写真画像データに、たとえばDublin Core、Image Regions、FOAFなどメタ情報を付与する。同じく、コンピュータはナレーションにもアノテーションを付与し、たとえば記憶手段(22)に格納するこの記憶手段に格納している写真画像データを用いてビデオコンテンツを作成するが、コンピュータ(12)は、各写真に付与する映像効果要素、ナレーション要素および反応時間要素の局所的制約をたとえばタイムライン上に設定する。レンダリング処理(S5)における時間調整手段は、たとえばコンピュータであって、各写真に対する映像効果要素、ナレーション要素および反応時間要素の各所要時間の総和を計算し、それが、たとえばビデオコンテンツの所望再生時間になるように調整する。

30

40

【 0 0 1 4 】

請求項1の発明では、写真のアノテーションされた領域に対するケンパーンズ効果などの映像効果要素と、オーディオの効果で特にナレーション要素との間のセマンティックな同期制約を考慮することで、簡単に魅力的なスライドショービデオ(ビデオコンテンツ)を製作できる。しかも、視聴者の反応時間も確実に確保できるので、視聴者による反応中に次のナレーションがオーバーラップするなどの不具合も生じない。

【 0 0 1 5 】

請求項2の発明は、写真遷移にフェードアウト/フェードインの視覚効果を用い、同じ写真内での人から人へ遷移にパンズームの視覚効果を用い、写真の再生期間中にナレーションおよび反応時間を配置したビデオコンテンツを作成する方法であって、最小フェード

50

アウト/フェードイン時間、ナレーション時間、反応時間および最小パンズーム時間の総和より最大ビデオ再生時間が小さいとき写真を除外し、そして最大フェードアウト/フェードイン時間、ナレーション時間、反応時間および最大パンズーム時間の総和より最小ビデオ再生時間が大きいとき、写真を追加するようにした、ビデオコンテンツ作成方法である。

【発明の効果】

【0018】

この発明によれば、写真のアノテーションされた領域に対するケンパース効果などの映像効果要素と、ナレーション要素および反応時間要素との間のセマンティック (semantic) な同期制約を考慮することで、簡単に魅力的なビデオコンテンツを製作できる。

10

【0019】

この発明の上述の目的、その他の目的、特徴および利点は、図面を参照して行う以下の実施例の詳細な説明から一層明らかとなる。

【発明を実施するための最良の形態】

【0020】

図1を参照して、この実施例のビデオコンテンツ作成装置10は、コンピュータ12を含む。このコンピュータ12としては、パーソナルコンピュータやワークステーションが利用可能である。

【0021】

コンピュータ12にはハードディスクやRAMのような内部メモリ14が設けられていて、その内部メモリ14には、メタ情報を入力するためのツールとして、たとえば、DublinCore (<http://dublincore.org>)、ImageRegions (<http://www.w3.org>)、FOAF (<http://www.foaf-project.org>)、Jena2(2.1) (<http://jena.sourceforge.net>)などが設定されている。これらはいずれも、写真画像データに関連するメタ情報を入力または登録もしくは付与する手段として機能する。

20

【0022】

ここで、メタ情報とは、データに関する構造化した情報を意味し、実施例の場合、取り扱う写真画像データの詳細を構造的に記述する情報である。このようなメタ情報を付与することをアノテーション (annotation) という。

【0023】

Dublin Coreはメタ情報を取り扱う代表的なツールとして知られていて、さらに、写真中の人物の顔やオブジェクトの領域 (以後、「リージョン」と呼ぶ。) を複数個指定し、これをメタ情報として保持するためにImage Regionsを利用する。リージョンが人の場合には、FOAFを利用して人のメタ情報を付与する。リージョンは静止画である写真に対して、そこへのズーム、リージョン間のパンといったエフェクトを加えるために用いる。メタ情報をRDFの形式でデータベースに格納するために、Jena2(2.1)を使用する。

30

【0024】

コンピュータ12は図示しないが、グラフィックボードまたはプロセッサやサウンドボードまたはプロセッサを内蔵していて、それらを通して、GUI画面や他のグラフィクス (映像) がモニタ16に表示されるとともに、スピーカ18からBGM (バックグラウンドミュージック) などの音声が出力される。

40

【0025】

コンピュータ12にはさらに、写真データ入力装置20が接続される。この写真データ入力装置20は、イメージスキャナ、デジタルカメラ (デジカメ)、インターネット (Web) などの少なくとも1つを含む。イメージスキャナは、認知症者の過去の写真をスキャンしてカラーまたはモノクロの写真画像データを入力する。デジカメはリアルタイムで撮影した写真画像データを入力できる他、過去の写真を撮影してそれらの写真画像データを入力するという使い方もできる。インターネットは、遠隔地から送信される認知症者の過去の写真の写真画像データを入力し、さらには必要に応じて認知症者の過去にまつわる事象の写真画像データを取り込むために使用できる。さらに他の種類の写真データ入力

50

装置が用いられてもよい。

【0026】

さらに、コンピュータ12には、インタフェース24を介してデータベース22が結合される。この実施例では、データベース22にはPostgreSQL7.4 (<http://www.postgresql.org>)というリレーショナルデータベースを用いる。

【0027】

なお、図示していないが、コンピュータ12は当然、キーボードやマウスなどの入力手段を持っている。

【0028】

図1の実施例のビデオコンテンツ作成装置10を用いて思い出ビデオを作成するためには、コンピュータ12などは図2に示す手順で動作する。

10

【0029】

まず、ステップS1で、思い出ビデオに使用する可能性のある写真画像データを入力するとともに、各写真に対してアノテーションを付与する。具体的には、図1の写真データ入力装置20を用いて、主として、当該認知症者の過去の写真の写真画像データをコンピュータ12に入力する。

【0030】

このとき、図3に示すGUI (Graphical User Interface) 画面26がモニタ16に表示される。このGUI26は、モニタ画面の左方の大部分を占める、写真表示編集領域28を含む。この写真表示編集領域28は、そこに入力した写真画像データの写真を表示し、さらには、リージョンを指定するなどの編集作業のために利用される。GUI26は、モニタ画面の右方に形成されるサムネイル表示領域30を含む。サムネイル表示領域30には、入力しかつ後述のようにして検索した写真画像データから使用する写真を選択するためにサムネイル画像を表示する。

20

【0031】

GUI26には、モニタ画面の下方に形成される、第1メタ情報入力領域32、モニタ画面のほぼ中央に形成される第2メタ情報入力領域34、およびモニタ画面の右下方に形成される再生順設定領域36が設けられる。メタ情報入力領域32には、領域28に表示もされている写真全般に関わるDublin Coreで定義されたメタ情報を入力する。また、メタ情報入力領域34はポップアップ形式で入力時に現れる領域であり、それを用いて、FOAFを利用してリージョンが人の場合のメタ情報を入力する。そして、再生順設定領域36は、思い出ビデオに取り込んだ写真を再生する順番を設定するために利用され、後に説明するように、ドラッグアンドドロップで順番を入れ替えることができる。

30

【0032】

ステップS1での写真入力およびメタ情報登録について、具体的に説明する。図4を参照し、今、写真データ入力装置20から図4に示すような写真(図面では線画であるが、実際は写真である。他の図面でも同様。)の写真画像データが入力されると、コンピュータ12は、モニタ16のGUI26の写真表示編集領域28にその画像データで表される写真(静止画)を表示する。それとともに、コンピュータ12は、その写真のサムネイル画像をサムネイル表示領域30に表示する。

40

【0033】

そして、図示しないマウス等を利用して、その写真中で2つのリージョン、リージョン1およびリージョン2を設定すると、図5に示すように、写真表示編集領域28中に、リージョン1およびリージョン2をそれぞれ特定する矩形枠29aおよび29bが表示される。このようなリージョン1およびリージョン2を指定すると、Image Regionsで定義される形式で、図6に示す各リージョンのメタ情報が設定される。つまり、リージョン1の原点(矩形枠の左上の角)の座標(x11、y11)およびその対角の座標(x12、y12)が登録され、さらにそのリージョン1の高さh1および幅w1もメタ情報として登録される。同様に、リージョン2についても、原点座標、対角座標、高さ、および幅が、それぞれ、x21、y21、x22、y22、h2、およびw2として登録される。

50

【 0 0 3 4 】

このようにしてステップ S 1 で写真を入力し、メタ情報を登録すると、たとえば図 7 に示すようなデータベースができる。この図 7 では右に、図 5 に示した、実際の写真が表示されリージョンが指定されている写真表示編集領域 2 8 が描かれている。そして、楕円形の中に「dc:」とあり、それに関連する矩形の中のデータが、Dublin Coreで登録したメタ情報である。たとえば「dc:date」では日付「20040716(2004年7月16日)」が、「dc:title」では名称「at Disney Animal Kingdom(ディズニー動物王国にて)」が、「dc:description」では説明文「They are very happy.(彼等は非常に楽しそう)」がメタ情報として登録される。

【 0 0 3 5 】

楕円形のなかに「imgReg:」とあるメタ情報は、Image Regionsでリージョンを指定したときのメタ情報である。「imgReg:has region」はリージョンが設定されていることを示すメタ情報であり、「imgReg:Rectangle」はリージョンが矩形であることを示し、「imgReg:regionDepict」はリージョンの説明で、「imgReg:boundingBox」はリージョンの原点位置とサイズ(高さh、幅w)とを含む。「imgReg:coords」はリージョンの原点位置および対角位置の座標である。

【 0 0 3 6 】

また、「foaf:gender」で与えられるメタ情報は、リージョンが人である場合の性別(例示では「female(女)»)であり、「foaf:name」で与えられるメタ情報は名前(例示では「Haruka(はるか)»)であり、「foaf:Person」で与えられるメタ情報はViewer(この思い出ビデオを観る人)と写真に写った人との関係を示し、例示では、「孫(grandchild)」であることがわかる。

【 0 0 3 7 】

なお、図 7 に示す各メタ情報は、それぞれ以下に示される。

【 0 0 3 8 】

```

x m l n s : i m g R e g = " http://www.w3.org/2004/02/image-regions# "
x m l n s : f o a f = " xmlns.com/foaf/0.1 "
x m l n s : d c = " http://purl.org/dc/elements/1.1/ "

```

このようにして、図 2 のステップ S 1 で写真入力と、メタ情報登録が行なわれると、コンピュータ 1 2 は、その写真と、それに付与したメタ情報とをデータベース 2 2 に格納する。なお、上の説明では 1 枚の写真とそれのメタ情報について説明したが、入力装置 2 0 で写真画像データを入力する都度、同じようにして、図 7 に示すようなメタ情報が登録され、そのメタ情報を付与した写真データがデータベース 2 2 に格納される。

【 0 0 3 9 】

ステップ S 1 では、さらに、さらにナレーションについてもアノテーションを付与してデータベース化した。つまり、前にも述べたように、思い出ビデオにとってナレーションが重要であるが、ナレーション付与は、非常に手間がかかる作業である。したがって、この実施例では、ナレーション付与作業の時間短縮のため、数千の典型的なナレーションテキストと、それらに対応する音声データとを含む、ナレーションデータベースを構築し、データベース 2 2 (図 1)に登録した。

【 0 0 4 0 】

後に詳細に説明するように、典型的なナレーションテキストは、日本語の文法に基づいた簡単な規則を使用して作成した。典型的なナレーションテキストは、その規則に対して、副詞、形容詞、および幾つかの名詞の組み合わせを与えて生成した。また、これらの語は、ナレーションのアノテーションに使用し、これらの語の組み合わせは共起確率に基づいて決定した。そして、音声合成技術を使って、またはプロのナレーターがそれらを読み上げて音声データを作成し、それをデータベース化し、同じくデータベース 2 2 に登録した。

【 0 0 4 1 】

このとき、ビデオ製作者が、何千ものナレーションデータから、全ての写真の、全ての

10

20

30

40

50

写真中の人物に対して、最も適切なナレーションを容易に選択できるようにするために、この実施例では、ナレーションと写真の両方のアノテーションを関連付けてナレーションデータを絞り込むようにしたので、ビデオ製作者は、絞り込まれたナレーションから適切なものを選択するだけで、ナレーションデータの設定が可能である。

【 0 0 4 2 】

図 8 はナレーションに付与されたアノテーションの例である。ナレーションのアノテーションには独自のボキャブラリーを定義した。図 8 で na はこのボキャブラリーのネームスペースを表す。" Narration1.wav " のナレーションテキストは、na:text によって指定される。そして " Narration1.wav " は na:keyword によって指定される、幾つかのインスタンスを指す。図中の左のインスタンスはナレーションテキストに関連する人を示す。それは na:referTo により foaf:Person のインスタンスを指している。一方、右側のインスタンスは、その人がどのように見えるかを示しており、dc:description が当該のナレーションテキストを作る際に使用した語を示している。

10

【 0 0 4 3 】

その後、図 2 のステップ S 2 では、ステップ S 1 で入力した写真のうち、そのとき使う写真を検索し、使用する写真の組を選択する。ただし、写真の検索では、上で説明したメタ情報が利用される。メタ情報を利用した写真の検索条件としては、たとえば、以下のものが利用できる。

【 0 0 4 4 】

まず、FOAF ではリージョンの種類が人であるとき、その人を特定するメタ情報を付与するのであるから、この FOAF のメタ情報を利用して、「特定の人物の写っている写真」を検索することができる。複数の人物を同時に検索できるが、この場合には、検索した名前の全員が写っている写真が対象となる。

20

【 0 0 4 5 】

Dublin Core のメタ情報を利用する場合には、「撮影年月日」で検索できる。たとえば、「From (第 1 指定日) ~ To (第 2 指定日)」で第 1 指定日以降第 2 指定日以前に撮影した全ての写真が検索できる。「From (指定日)」でその指定日以降に撮影した全ての写真が検索できる。同様に、「To (指定日)」でその指定日以前に撮影した全ての写真が検索できる。また、「特定のプロパティに特定の値が含まれている写真」を検索できる。たとえば、「dc:title」に「Disney」の文字が含まれる写真など。ただし、複数同時に選択または検索可能であるが、複数選択時には、設定された検索条件を同時に充足する写真だけが対象となる。

30

【 0 0 4 6 】

そして、検索した写真から実際に使用する写真の組を選択する場合、その組の複数の写真の再生順序を指定または決定する。再生順を設定するためには、たとえば、図 9 のような GUI 2 6 において、サムネイル表示領域 3 0 に表示されるので、その一覧画像の中の写真を再生順序設定領域 3 6 にドラッグアンドドロップすることによって、選択した写真の再生順を設定することができる。ただし、ここでは、その具体的な詳細は省略する。

【 0 0 4 7 】

このステップ S 2 で使用する写真の組を選択したとき、好ましくは、それらの写真の各々に優先順位を付与しておく、この優先順位は、思い出ビデオに使用したい順位のことである。たとえば後に説明するように、思い出ビデオ全体の総時間数がたとえば 2 0 分とか 3 0 分とか限られているとき、映像効果やナレーションに必要な時間を計算すると写真の枚数を減らさざるを得ないとき、最も低い優先順位の写真から除いていく作業をする。

40

【 0 0 4 8 】

つづいて、図 2 のステップ S 3 で、写真の再生時に流す BGM およびナレーションを選択する。この実施例では、たとえば MP3 形式の楽曲ファイルをサポートしていて、その中から BGM として使う楽曲をプレイリストとして登録すれば、BGM がそのプレイリストに沿って、再生されるようにしているが、BGM についてはここではこれ以上の説明は省略する。

50

【 0 0 4 9 】

次に図2のステップS4では、プレビューをするかどうか判断する。写真の選択などをすべて終えたときにこのステップS4で“YES”が判断されることになるが、そうでないときには、“NO”と判断され、ステップS2やステップS3に戻って作業を続ける。つまり、図2の各ステップS1-S4はそれぞれ、メニューでの選択によって、任意の時間に任意の作業量で何回も実行可能であるので、製作者は、時間のあるときに、必要なだけ必要な作業を行えばよい。いずれの場合にも、前回までの作業の結果はデータベース22に格納されているので、今回の作業では、まず、データベース22から前回までのデータを読み出し、その後それに続行した処理またはそれを変更する処理を行う。

【 0 0 5 0 】

そして、ステップS4で“YES”を判断したときには、続くステップS5で、セマンティクスを考慮したメディア同期の手法に則り、ケンパーズ効果、BGMおよびナレーションを付与することによって、たとえばRVMML形式の思い出ビデオをレンダリングする。

【 0 0 5 1 】

なお、「RVMML」とは、swfを完全に表現できるように設計されたXMLの一種であり、すべてのバージョンのswf動画はRVMMLとして表現できる。ただし、swfは、Flashのバージョンやフレームレートなどの情報を持つファイルヘッダを除くと、基本的にはタグの列だけである。たとえば、定義タグで図形を定義し、操作タグでその図形をフレームに配置し、表示タグで現在のフレームを画面に描画する、というパターンが1フレームに相当し、これを繰り返す。

【 0 0 5 2 】

そして、生成されたRVMMLは、RVMMLからFlashツールであるKineticFusion (<http://www.kinesoftware.com>)を使用すれば、Flashムービー形式の思い出ビデオが作成できる。

【 0 0 5 3 】

図2のステップS5は、具体的には、図10に示す手順で実行される。最初のステップS11では、コンピュータ12は、まず、BGMとしてのプレイリスト(図示せず)の楽曲の再生を開始する。以後、楽曲は、そのプレイリストに沿って、順次切り替わり、かつ連続的に再生される。つまり、コンピュータ12は、プレイリストに登録した楽曲データを読み出し、それをサウンドボードまたはプロセサで処理させることによって、スピーカ18(図1)から、当該プレイリストの楽曲が音声として再生される。

【 0 0 5 4 】

次のステップS13では、コンピュータ12は、モニタ16の表示画面に形成されたGUI26の写真表示編集領域28(図3)の幅をSw、高さをShとして設定する。

【 0 0 5 5 】

ついで、ステップS15で、コンピュータ12は、写真個数nをインクリメント($n = n + 1$)する。そして、次のステップS17では、再生順設定領域36(図9)に挙げられたn番目の写真がフェードイン態様で、表示編集領域28に表示される。つまり、コンピュータ12は、データベース22(図1)から再生順のn番目の写真とそれに付随するメタ情報(アノテーション)とを読み出し、n番目の写真を表示する。ただし、最初は $n = 1$ であるので、1番目の写真がフェードイン表示される。

【 0 0 5 6 】

そして、次のステップS18において、コンピュータ12は、その写真についてナレーションが付与されているかどうか判断し、既にナレーションが付与されて登録されている場合には、このナレーションをステップS19で再生する。したがって、このステップS19において、既登録のナレーションが確認できる。

【 0 0 5 7 】

ただし、ナレーションが未だ付与されていないときには、その後、ステップS20での一定時間のポーズの後、次のステップS21で、コンピュータ12は、そのn番目の写真

10

20

30

40

50

にリージョンが指定されているかどうか、たとえばImage Regionsのメタ情報があるかで判断する。リージョンが指定されているなら、次のステップS 2 3で、コンピュータ1 2は、リージョン番号mをインクリメント($m + 1$)する。そして、ステップS 2 5で、m番目のリージョンのImage Regionsのメタ情報を参照する。このメタ情報に、m番目のリージョンの位置データやサイズデータが含まれる。したがって、次のステップS 2 7で、コンピュータ1 2は、そのようなメタ情報および先に設定した表示領域の高さShおよび幅Swを利用して、m番目のリージョンを、領域2 8の中央に位置合わせする。

【0058】

一例として、そのリージョンの幅を w_1 、高さを h_1 とすると、拡大率を $\min[Sw/w_1, Sh/h_1]$ として、そのリージョンがちょうど表示画面2 8の中央に収まるまで、1フレームずつ、表示画面2 8に対して写真画像を横軸にdX、縦軸にdY移動し、dZ分拡大して表示する。

10

【0059】

ただし、このステップS 2 7では、そのリージョンの画像を表示領域の中央に位置合わせするだけでなく、たとえば、左上、右下など他の位置に位置合わせするようにしてもよい。

【0060】

その後、ステップS 2 9でのポーズの後、次のステップS 3 1で、コンピュータ1 2は、残りリージョンがなくなったかどうかを判断する。つまり、Image Regionsのメタ情報からリージョン個数があるので、このステップS 3 1では、ステップS 2 3でインクリメントした結果がそのリージョン個数に等しくなったかどうか判断すればよい。

20

【0061】

残りリージョンがあれば、次のステップS 3 2において、コンピュータ1 2は、表1に示す識別子RT - 1、RT - 2およびRT - 4で示される映像効果A - 1、A - 2を付与する。

【0062】

ここで、映像効果について説明する。発明者等は、映像編集の経験のある複数のクリエイター(製作者)に、写真から思い出ビデオに変換する際に付加できる映像効果(各写真に付与したエフェクト、写真間のトランジション)について、どのようなものが想定できるか、聞き取り調査した。その結果は、以下のA - 1 ~ A - 3およびB - 1 ~ B - 5であった

30

(A) エフェクト

A - 1 : 写真中の人物の顔を含む矩形領域(以下、リージョンと呼ぶ)に対するズームアップ、パンという、いわゆるケンバーズ(Ken Burns)効果。

A - 2 : パンの代わりに、ズームアップされたリージョンをフェードアウトして、次のリージョンをフェードインする。

A - 3 : カラー写真をまずモノクロ調に表示した後、徐々にカラー表示に遷移させていく。

(B) トランジション

B - 1 : 前の写真をフェードアウトしながら次の写真をフェードインし、それらをオーバーラップさせる。

40

B - 2 : 次の写真をスライドインさせる。

B - 3 : 前の写真をディゾルブして次の写真に遷移する。

B - 4 : 前の写真をページピール(右下隅から捲り上げるような効果)して次の写真に遷移する。

B - 5 : 前の写真を中心の縦軸で回転させて、次の写真に遷移する。

【0063】

上記の結果を踏まえて、表1に示すような思い出ビデオをレンダリングするためのテンプレートを作成した。表1の「使用された情報」に挙げた項目を写真に付与すべきアノテーションとし、アノテーションオントロジ(ontology)として設計した。アノテーション

50

付与の枠組みとして、セマンティックWeb（これは、検索性能の向上や利便性を高める次世代Web技術で、「メタ情報（内容を説明する付加情報）」と「オントロジ（メタデータを記述する用語の定義）」という2つの技術を用いる。）の枠組みを用いた。すなわち、アノテーションは、RDF（Resource Description Framework）で記述される。これは、将来的には他人がアノテーションした写真を自分や家族の思い出ビデオに利用するような情報交換を考えたとき、Webとの親和性を考慮したためである。

【0064】

【表1】

識別子	エフェクト、トランジションが使用される条件	エフェクト、トランジション	使用された情報
RT-1	リージョン間のパンの上下移動量／左右移動量が閾値以下のとき	A-1	写真中の人物領域
RT-2	リージョン間のパンの上下移動量／左右移動量が閾値以上のとき	A-2	
RT-3	写真がモノクロからカラーに変わるとき	A-3	写真の色調
RT-4	視聴者に思い出深い被写体に対応するリージョンにだけエフェクトを適用する	A-1 A-2	写真中の人物情報
RT-5	写真間の年代が閾値以下のとき	B-1	撮影された日付
RT-6	写真間の年代が閾値以上のとき	B-5	
RT-7	シナリオが一貫しているとき	B-1	撮影された日付 撮影された出来事
RT-8	使用される写真の年代間隔が閾値以下の場合に、シナリオの転換点のとき	B-5	

【0065】

さらに、発明者等は、既存のボキャブラリーを可能な限り利用するという方針をたて、図1実施例のための実験では、撮影された日付や出来事に関しては、書誌情報に関する標準的なボキャブラリーである上述のDublin Coreで記述した。また、写真中の人物の情報を記述するためには、人の情報を記述するための標準的なボキャブラリーである上述のFOAFを用いた。そして、写真中の人物領域を記述するためには、上述のImage Regionを使用し、写真の色調はExif (<http://it.jeita.or.jp/document/publica/standard/exif/english/jeida49e.htm>参照)の色空間情報から取得できる。また、写真中の人物(被写体)との関係については、FOAFでknowsプロパティが定義されているが、思い出ビデオ作成においては本人(視聴者)と被写体(関係者)との間柄をより詳細に定義する必要があるため、FOAFのknowsの属性を拡張して定義されたRELATIONSHIP (<http://vocab.org/relationship/>参照)を利用した。これを用いて、親子関係、親戚関係などを記述した。そのサンプルが先の図7に示される。

【0066】

さらに、表1に挙げたエフェクト、トランジションに対して、思い出ビデオ作成に用いる映像効果のオントロジを表2のように定義した。

【0067】

【表 2】

エフェクト, トランジション	レンダリング オントロジ	説明
A-1	Zoom	リージョンへのズーム
	Pan	リージョン間のパン
A-2	RegionOverlap Transition	リージョン間のフェードイン, フェードアウト
A-3	Mono2Color	モノクロからカラーへ
B-4	PhotoOverlap Transition	フェードイン, フェードアウト による写真の遷移
B-5	RotateTransition	回転による写真の遷移

10

【 0 0 6 8 】

そして、写真のアノテーションから映像効果への変換ルールを記述し、テンプレートとして使用する。表 3 に、表 1 の「RT - 8」のレンダリングテンプレートの表現例を示す。ただし、表 1 に示す他の識別子に付いても同様にテンプレートを作成しておくものとする。

20

【 0 0 6 9 】

【表 3】

```

...
<rt:RotateTransition rdf:ID="RT-8">
  <rt:photoInterval rdf:datatype="&xsd;unsignedLong">
    2592000      例えば日付の範囲が1ヶ月(秒)
  </rt:photoInterval>
  <rt:scenarioContinuity rdf:datatype="&xsd;boolean">
    true        シナリオの連続性が真
  </rt:scenarioContinuity >
</rt:RotateTransition>
...
SELECT ?transition
WHERE (<firstPhotoURI>, dc:date, ?data_1) 最初の写真の日付
      (<lastPhotoURI>, dc:date, ?data_2)   最後の写真の日付
      (<curPhotoURI>, dc:subject, ?scenario_1) 現在の写真のテーマ
      (<nextPhotoURI>, dc:subject, ?scenario_2) 次の写真のテーマ
      (?transition, rt:pPhotoInterval, ?interval) 制約を満たす
      (?transition, rt:scenarioContinuity, ?scenario) トランジションの検索
AND   (?interval >= (?data_2 - ?data_1)) 日付の範囲の制約
AND   (?scenario != (?scenario_1 EQ ?scenario_2))シナリオ連続性の制約
USING
      dc FOR <http://purl.org/dc/elements/1.1> レンダリングオントロジ定義
      rt FOR <http://therapy.atr.jp/vocab/rendering_technique#>

```

RDFで記述されたレンダリング
オントロジのテーブルの一部 (RT-8)

写真のアノテーションオントロジとレンダ
リングオントロジの変換ルール (RT-8)

30

40

【 0 0 7 0 】

ここでは、レンダリングテンプレート自身も R D F のステートメントとして記述し、上述の R D Q L を用いて、レンダリングテンプレートを検索することを想定している。なお、レンダリングオントロジについては、インタオペラビリティ（相互操作性）の必要性は低いいため、別の独自形式を定義してもよい。

50

【 0 0 7 1 】

ステップ S 3 2 に戻って、具体的には、コンピュータ 1 2 は、図 7 のようなメタ情報のうち、リージョン座標値 (imgReg:coords) およびリージョン境界値 (imgReg:boundingBox) から、複数のリージョン間間隔、X 方向 (横方向) 間隔および Y 方向 (縦方向) 間隔を算出する。そして、それらのリージョン間間隔が所定の閾値以上か以下かを判別する。そして、X 方向 (横方向) 間隔および Y 方向 (縦方向) 間隔のいずれかが閾値以下のときには、ケンバーズ効果を使用するべく、写真中の人物の顔を含むリージョンに対するズームアップをした後、次のリージョンへパンさせる。逆に、X 方向 (横方向) 間隔および Y 方向 (縦方向) 間隔のいずれかが閾値以上のときには、パンの代わりに、ズームアップされたリージョンをフェードアウトして、次のリージョンをフェードインさせる。

10

【 0 0 7 2 】

また、図 7 に示す foaf:person で示す被写体情報が、視聴者に思い出深い被写体であると、その被写体のリージョンだけにズームアップをし、その後、リージョン間間隔に応じたパンかフェードアウト、フェードインからの処理を行う。

【 0 0 7 3 】

ただし、このステップ S 3 2 において識別子 RT - 1、- 2 および - 4 のような映像効果を A - 1 および A - 2 を自動的に付与するためには、表 3 で例示した変換テンプレートが用いられる。

【 0 0 7 4 】

その後、先のステップ S 2 3 に戻り、ステップ S 2 3 - S 3 2 を繰り返し実行する。

20

【 0 0 7 5 】

ステップ S 3 1 で “ NO ” なら、つまり、写真リスト欄 3 8 で n 番目のその写真の全てのリージョンの処理が終わったなら、次のステップ S 3 3 で、コンピュータ 1 2 は、n 番目の写真を写真表示編集領域 2 8 からフェードアウトさせる。

【 0 0 7 6 】

続いて、ステップ S 3 5 において、コンピュータ 1 2 は、処理すべき写真の残りがなくなったかどうかを判断する。つまり、図 9 で示される写真リスト中の写真の数は予め判っているので、このステップ S 3 5 では、ステップ S 1 5 でインクリメントした結果がその枚数に等しくなったかどうか判断すればよい。

【 0 0 7 7 】

写真が未だ残っていれば、次のステップ S 3 6 において、コンピュータ 1 2 は、表 1 に示す識別子 RT - 3、RT - 5、RT - 6、RT - 7 および RT - 8 で示される映像効果 A - 3 や B - 1 および B - 5 を付与する。

30

【 0 0 7 8 】

具体的には、コンピュータ 1 2 は、メタ情報のうち、図 7 には図示していない、写真の色調のメタ情報を前述の Exif の色空間情報から取得し、その色情報が、前の写真がモノクロで今回の写真がカラーのような色変化を示しているかどうか判断する。色変化情報があれば、コンピュータ 1 2 は、今回のカラー写真をまずモノクロ調に表示した後、徐々にカラー表示に遷移させていくという映像効果を付与する。

【 0 0 7 9 】

また、コンピュータ 1 2 は、図 7 に示すメタ情報のうち、撮影日 (date) のデータを取得し、前の写真と今回の写真との間の年代差を算出する。そして、その年代差が所定の閾値以上か以下かを判別する。前の写真との間の年代差が閾値以下のときには、コンピュータ 1 2 は、前の写真をフェードアウトしながら次の写真をフェードインし、それらをオーバーラップさせる、という映像効果を付与する。前の写真との間の年代差が閾値以上のときには、コンピュータ 1 2 は、前の写真を中心の縦軸で回転させて、今回の写真に遷移する、という映像効果を付与する。

40

【 0 0 8 0 】

さらに、コンピュータ 1 2 は、図 7 に示すメタ情報のうち、撮影日や出来事 (title) を参照して、シナリオが一貫しているかどうか、判断する。これは、たとえば、写真が年

50

代順に取り込まれているか、とか同じ年代であれば季節の順序に従っているかなどを判断すればよい。そして、シナリオが一貫していると判断したときには、前の写真をフェードアウトしながら次の写真をフェードインし、それらをオーバーラップさせる、という映像効果を付与する。

【0081】

さらに、上述のようにして算出した写真の年代差が所定の閾値以下ではあるが、かつシナリオの転換点であるような場合、たとえば、進学した、結婚した、子供が生まれた、などのような場合には、コンピュータ12は、前の写真を中心の縦軸で回転させて、今回の写真に遷移する、という映像効果を付与する。

【0082】

ただし、このステップS36での映像効果を自動的に付与するためには、表3で例示した変換テンプレートが用いられる。

【0083】

なお、先のステップS35で写真残数があると判断したときには、先のステップS15に戻り、ステップS15 - S36を繰り返し実行する。ステップS35で“NO”なら、コンピュータ12はステップS37でBGMを停止し、終了する。

【0084】

このようにして、図10のフロー図に従って、写真画像データをそれに関連するメタ情報を用いて編集しさらには映像効果付与することによって、一連のビデオコンテンツ(思い出ビデオ)が生成(レンダリング)されるが、この映像効果付与と同時並行して、またはそれに続いて、ナレーション付与を行う。

【0085】

ナレーション付与を実行するに際しては、図4に示すGUI26に形成されるナレーションボタン38を操作する。ユーザは、生成したスライドショーにナレーションを付与する必要があるとき、マウス(図示せず)でこのナレーションボタン38をクリックする。そうすると、図11に示すように、GUI26のほぼ中央に、写真表示編集領域28その他にオーバーラップする形で、ナレーション作成領域44が、ポップアップ形式で表示される。

【0086】

このナレーション作成領域44には、ナレーション候補テキスト表示領域46が形成される。ナレーション候補テキスト表示領域46は、後に説明するように、メタ情報から取得したキーワードもしくはユーザが入力したキーワードに基づいてコンピュータ12がナレーション候補テキストを作成したとき、そのナレーション候補テキスト(文)を表示するための領域である。このナレーション候補テキスト表示領域46に関連して、OKボタン48が配置される。このOKボタン48は、ナレーション候補テキスト表示領域46に表示されたナレーションテキスト文を選択するかどうかを指示するためのものであり、表示されているナレーション候補を使用するならそれをクリックすればよい。

【0087】

ただし、ナレーション候補テキスト表示領域46には、1つだけの候補ではなく、複数の候補が一度に表示されるようにしてもよく、その場合には、ユーザは、その領域46を直接クリックして1つまたは複数のナレーションテキストを選択するようにすればよい。

【0088】

ナレーション候補テキスト表示領域46の下方には、指示ボタン50が形成される。たとえば、コンピュータ12が適当なナレーション候補テキストを作成できなかったような場合に、ユーザが適当なキーワードを手動で入力するときにこの指示ボタン50をマウスでクリックする。そうすると、その下のキーワード入力領域52が有効化される。したがって、ユーザは図示しないキーボードを使用してキーワードを入力することができる。このキーワード入力領域52は、好ましくは、図11に示すように、カテゴリ毎にキーワードを入力できるように、複数(この実施例では5つ)のカテゴリに区分けされている。この入力領域52を見れば、ユーザは、意図したキーワードが入力されたかどうか

10

20

30

40

50

を確認することができる。そして、意図どおりのキーワードを入力していることを確認したとき、ユーザは、入力領域 5 2 の右にある OK ボタン 5 4 を操作すればよい。それによって、キーワード入力が終了できる。

【 0 0 8 9 】

キーワード入力 OK ボタン 7 1 の右には、終了ボタン 5 6 が設定される。この終了ボタン 7 2 は、ナレーション付与作業を終了するとき操作する。

【 0 0 9 0 】

このようにして、ナレーション付与ボタン 3 8 を操作することによって、GUI 2 6 が図 1 1 に示す状態となり、その状態で、図 1 2 に示すナレーション付与動作を実行することができる。

10

【 0 0 9 1 】

図 1 2 の最初のステップ S 4 1 で、コンピュータ 1 2 は、ナレーションを付与すべき写真が選択されたかどうか判断する。ユーザが写真を選択するためには、先に説明したように、サムネイル表示領域 3 0 に表示されたサムネイル画像をクリックすればよい。そして、コンピュータ 1 2 は、サムネイル表示領域 3 0 での操作によってステップ S 4 1 で写真が選択されたと判断すると、次のステップ S 4 3 において、選択したサムネイル画像で表される写真を、写真表示編集領域 2 8 に表示する。ただし、このとき GUI 2 6 はナレーション作成領域 4 4 がその写真表示編集領域 2 8 にオーバーラップした「ナレーション付与モード」になっているので、この領域 2 8 で写真の編集をすることはできない。

【 0 0 9 2 】

20

次のステップ S 4 5 において、コンピュータ 1 2 は、先に述べた方法で登録されているメタ情報を取得し、次のステップ S 4 7 で、そのメタ情報からキーワードを抽出する。そして、ステップ S 4 9 において、コンピュータ 1 2 は、そのキーワードに基づいて、ナレーション候補テキストをナレーション候補テキスト表示領域 4 6 に表示する。

【 0 0 9 3 】

ここで、ナレーション候補の生成方法について、説明する。

【 0 0 9 4 】

この実施例のシステム 1 0 では、コンピュータ 1 2 の内部メモリ（図示せず）またはデータベース 2 2（図 1）に、表 4 に示すような、典型的なナレーションパターンを予め設定しておく。その意味で、これら内部メモリおよび/またはデータベース 2 2 がナレーションパターン設定手段として機能する。そして、この表 4 に示す単語 X 1 X 5 を、メタ情報から抽出したキーワードに基づいて当てはめることによって、ナレーションテキストを自動生成する。

30

【 0 0 9 5 】

【表 4】

- ・ X 1 ですね（ですか？）
- ・ これは X 1 ですね（ですか？）
- ・ X 2 と X 3 しましたね（しましたか？）
- ・ X 4 で X 3 しましたね（しましたか？）
- ・ X 5 に X 3 しましたね（しましたか？）

40

【 0 0 9 6 】

ただし、単語 X 1 X 5 の与え方は、一例として、表 5 に従う。

【 0 0 9 7 】

【表5】

- ・ X1 ← {[([副詞] + 形容詞) または (名詞1 + の)] + 名詞2}
名詞2 ← 人、物、場所、時候
- ・ X2 ← {[([副詞] + 形容詞) または (名詞1 + の)] + 名詞2}
名詞2 ← 人
- ・ X3 ← {[([副詞] + 名詞2)}
- 名詞2 ← 行為
- ・ X4 ← {[([副詞] + 形容詞) または (名詞1 + の)] + 名詞2}
名詞2 ← 場所
- ・ X5 ← {[([副詞] + 形容詞) または (名詞1 + の)] + 名詞2}
名詞2 ← 場所、時候

10

【0098】

たとえば、表4の第1パターン「X1ですね(orですか)」や第2パターン「これはX1ですね(orですか)」を使うときには、単語X1には、人、物、場所、時候のいずれかである名詞2を当てはめる。ただし、その名詞2にも何種類があり、それが{([副詞] + 形容詞) または (名詞1 + の) + 名詞2}で表されている。これを分解すると、単語X1に当てはめられる単語は、「形容詞 + 名詞2」(たとえば、「きれいな」)、「副詞 + 形容詞 + 名詞2」(たとえば、「大変きれいな花」)、「形容詞 + 名詞1の名詞2」(たとえば、「きれいな庭の花」)、「副詞 + 形容詞 + 名詞1の名詞2」(たとえば、「大変きれいな庭の花」)、「そして「名詞1の名詞2」(たとえば、「庭の花」)のいずれかとなる。

20

【0099】

表4の第3パターン「X2とX3しましたね(orしましたか)」を使うときには、単語X2には、人である名詞2を当てはめる。ただし、そのX2の場合の名詞2は、第1パターンや第2パターンのときと同様に、{([副詞] + 形容詞) または (名詞1 + の) + 名詞2}で定義され得る。また、単語X3には、行為である名詞2を当てはめる。このときの名詞2は、{([副詞] + 名詞2)}で定義される。つまり、副詞が付いたか、付かない名詞2(行為)(たとえば、「楽しく旅行」または「旅行」)である。

30

【0100】

表4の第4パターン「X4でX3しましたね(orしましたか)」を使うときには、単語X4には、場所である名詞2を当てはめる。ただし、その名詞2には、上述の場合と同じく、{([副詞] + 形容詞) または (名詞1 + の) + 名詞2}で定義される。なお、単語X3については上述のとおりである。たとえば、場所としては、「遊園地」、「デパート」などが例示できる。

【0101】

表4の第5パターン「X5でX3しましたね(orしましたか)」を使うときには、単語X5には、時候および/または場所である名詞2を当てはめる。ただし、その名詞2には、上述の場合と同じく、{([副詞] + 形容詞) または (名詞1 + の) + 名詞2}で定義される。時候としては、春夏秋冬、何月などが例示できる。なお、単語X3については上述のとおりである。

40

【0102】

さらに、上記において、名詞2に与える単語は、一般名詞の概念構造から、ナレーションに使用するのに粒度として適当なものを選択して用いる。そして、副詞、形容詞、名詞1については、その名詞2と適合するものを割り当てる。ただし、この「適合」については、図示しないが、適合表を作成し、名詞2を決定した後、その適合表を参照して副詞、

50

形容詞、名詞 1 を採用する。

【 0 1 0 3 】

このような表 1 のナレーションパターンの X 1 - X 5 に表 2 のように定義できる単語を当てはめる訳であるが、実施例では、この「単語」を、写真に付加したメタ情報から取得しようとするものである。

【 0 1 0 4 】

図 1 4 および図 1 5 には、実施例において単語決定手段として機能するシソーラス辞書 5 8 および共起辞書 6 0 が図解される。このようなシソーラス辞書 5 8 や共起辞書 6 0 は、ともに、データディクショナリ(データ辞書)の一種であり、実施例では、図 1 に示すデータベース 2 2 またはコンピュータ 1 2 の内部メモリに予め設定されている。

10

【 0 1 0 5 】

シソーラス辞書 5 8 は、図 1 4 に示すように、表記の揺れや、関連語、同義語、類似語などを概念的に類似したキーワードのツリー構造または網構造の階層構造として示したものであり、カテゴリ毎にまとめられている。図 1 4 の例では、カテゴリとして、「場所」および「行為」が示されている。たとえば、「偕楽園」、「後楽園」、「兼六園」などは「公園」という概念でくくられ、いずれも固有名詞である「ディズニーランド」や「USJ」は、「遊園地」という概念に当てはめられ、これら「公園」および「遊園地」は、別の概念たとえば「デパート」とともに、カテゴリとしては「場所」に含まれることを示している。「行為」というカテゴリについても同様である。「旅行」はまとめて「旅行」で表し、「旅行」は、「散歩」、「おでかけ」などともに、「行為」のカテゴリに分類されている。先のナレーションパターンでの「名詞 2」として適当な他のカテゴリ「人」、「時候」などについても、同様にキーワードを集積している。

20

【 0 1 0 6 】

特定の単語と単語との関係、結びつきを共起関係といい、共起辞書 6 0 には、図 1 5 に示すように、各カテゴリたとえば「場所」および「行為」にそれぞれ含まれるキーワード間の共起関係が示されている。この共起辞書 7 6 によれば、一例として、場所の「デパート」というキーワードは行為のなかでは、「おでかけ」というキーワードとしか繋がらないが、「公園」といえば、「おでかけ」、「散歩」および「旅行」という複数のキーワードに強いつながりを持つことがわかる。

【 0 1 0 7 】

30

このようなツールを利用して、表 4 のナレーションパターンに単語、特に名詞 2 を適用することによって、ステップ S 4 9 において、ナレーション候補テキストを作成する。

【 0 1 0 8 】

ナレーション候補の具体例を説明する。図 7 に示すメタ情報には、日付(dc:date)として「2004.07.16」があり、タイトル(dc:title)として「ディズニー動物王国にて」があり、人(faof:person)として「rel:grandchildOf」がありさらに、名前(foaf:name)として「はるか」が含まれる。

【 0 1 0 9 】

まず、日付が「2004.7.16」であることから、シソーラス辞書 5 8 を参照すれば、図 1 4 には図示していないが、カテゴリ「時候」としては「夏」であることがわかる。 「ディズニー動物王国にて」というタイトルに含まれる「ディズニー」を図 1 4 のシソーラス辞書 5 8 で検索すると、それは、「遊園地」に包含され、その「遊園地」は「場所」のカテゴリに該当することがわかる。さらに、図 1 5 の共起辞書 6 0 を参照すると、「遊園地」は 3 つの行為「おでかけ」、「散歩」、「旅行」に共起関係を有することがわかる。さらに、「人」として「孫」があり、その名前が「はるか」であることがわかる。このようにして、メタ情報からキーワードを抽出し、そのキーワードから、ナレーションパターンに適用可能な「単語」を検索すると、時候が「夏」で、場所が「遊園地」で、行為が「おでかけ」、「散歩」または「旅行」で、人が「孫のはるか」であることがそれぞれわかる。

40

【 0 1 1 0 】

50

したがって、コンピュータ12は、ステップS49で、一例として表6に示すような3つのナレーションテキストを生成する。

【0111】

【表6】

人	場所	物	時候	行為	ナレーションテキスト
---	遊園地	---	夏	---	夏の遊園地ですね
はるか	遊園地	---	夏	旅行	孫のはるかとは遊園地に旅行しましたね
	遊園地		夏	旅行	夏の遊園地に旅行しましたね

10

【0112】

表6の第1の候補は、第1のナレーションパターンを選択して単語を適用したものであり、第2の候補は、第3のナレーションパターンを選択して単語を適用したもので、第3の候補は、第5のナレーションパターンを選択して単語を適用したものである。

【0113】

そして、ステップS49で、このようなナレーションテキストを、ナレーション候補として、図11のナレーション候補テキスト表示領域46に一度に、または順次表示する。

【0114】

そして、ステップS51でコンピュータ12は、そのようなナレーション候補テキストが選択されたかどうか、判断する。どれかのナレーション候補テキストが選択されると、次のステップS53で、コンピュータ12は、このナレーションテキストの音声データを取得する。

20

【0115】

この実施例では、音声モデルを図1のデータベース22またはコンピュータ12の内部メモリに登録しておき、ナレーションテキストを決定すると、その音声モデルを使って音声合成の手法で、ナレーション音声を作成する。ただし、音声データは、音声合成による他、たとえばプロのナレーターが発声した音声データを内部メモリやデータベース22内に収録しておき、その音声データを編集することによって、ナレーション音声を作成するようにしてもよい。

30

【0116】

そして、ステップS55において、コンピュータ12は、ステップS53で取得し、また作成したナレーション音声データを、ステップS41で選択した写真に紐付けして、データベース22に登録する。

【0117】

このようにして、1枚の写真についてのナレーション音声データが写真に付与されるが、次のステップS57で、ナレーション付与をしたい写真が未だあるかどうか判断し、まだナレーション付与を続行するときは終了ボタン56(図11)を押さないで、先のステップS41に戻るが、終了ボタン72を押したなら、このステップS57で“YES”となり、ナレーション付与ステップS7が終了する。

40

【0118】

ただし、先のステップS49で表示したナレーション候補をステップS51で選択しなかったときには、コンピュータ12は、次のステップS59でさらにナレーション候補があるかどうか判断し、もしあれば、次のステップS61でナレーション候補を更新して、再び、ナレーション候補テキスト表示領域46に表示し(ステップS49)、ユーザの選択を待つ。

【0119】

また、ステップS59で適当なナレーション候補がないと判断したときには、コンピュータ12は、ユーザによる変更キーワードを受け付ける。ユーザはキーワードを入力する

50

ときには、指示ボタン50(図11)を操作し、キーワード入力領域52にキーワードを入力する。このとき、入力するキーワードは、ナレーションパターン(表4)の名詞1および名詞2に対応する{人、物、場所、時候、行為}、副詞や形容詞に対応する{どんな}をそれぞれ入力するものとする。ただし、指定しないカテゴリや項目があってもよい。

【0120】

このようにして、ユーザがキーワードを入力した後は、コンピュータ12は、ステップS49で、上で説明したように、ユーザ入力キーワードから、シソーラス辞書58や共起辞書60を使って「単語」を決定し、その単語を表1のナレーションパターンに当てはめて、ナレーションテキストを生成し、表示する。以後、先に説明したように、ユーザは、その表示されたナレーション候補テキストを選択し、ナレーションテキストを決定する。

10

【0121】

このようにして決定したナレーションテキストの音声データを写真画像データとともにデータベース22に登録することによって、たとえばRVM形式のビデオムービーのようなビデオコンテンツを製作することができるが、この実施例では、付与すべき映像効果として、現在は2つのタイプの映像効果を対象としている。1つは写真間のトランジションの際に適用するフェードアウト/フェードインである。もう1つのタイプは写真の幾つかの領域に視聴者の注意を引き付けるための、パン、ズームの効果である。特に後者の映像効果は、適切なナレーションが伴なければならない。たとえば、「あなたの息子さんは、かわいいですね」というナレーションは、視聴者の息子を含む領域にズームアップしているタイミングで再生しなくてはならない。また、映像効果との関連以外に、適切なポーズ(反応時間)がナレーションの後に挿入される必要がある。すなわち、視聴者が、ナレーションに反応することができる十分な時間を与える必要がある。ポーズが短過ぎると、視聴者が何らかの発話をしようとするのを、次のナレーションで遮られるかもしれず、それは視聴者にとって非常にいらだたく感じられる。

20

【0122】

つまり、視聴者にとって興味あるスライドショービデオにするためには、視覚的效果やBGM、ナレーションなどのオーディオ効果を付与する必要があるが、その際には写真の内容に応じて適切に視覚的效果とオーディオ効果を同期させることが必要となる。たとえば、ウェブ上のマルチメディアコンテンツの標準的な表現形式として、Synchronized Multimedia Integration Language (SMIL)が良く知られている。この表現形式は、タイムライン上に各コンテンツが割付済みのマルチメディアコンテンツを記述することを対象としている。この実施例で問題となるのは、アノテーションされた写真のセットから魅力的なスライドショービデオを作成する際の視覚的效果とオーディオ効果の間の意味的なレベルでの同期関係である。具体的には、ナレーションやBGMと、写真への視覚的效果(映像効果)を付与するタイミングを同期させる必要があり、そのことをここでは、「セマンティクス(semantics)を考慮したメディア同期」と呼ぶことにする。

30

【0123】

このような「セマンティクスを考慮したメディア同期」をとるために、この実施例では、上述の全ての要素(ナレーション、ポーズ、映像効果、写真および写真中の人物)を揃えた後に、映像効果要素、ナレーション要素およびポーズ(反応時間)要素についてのセマンティックな制約を考慮に入れて、ビデオのタイムライン上に並べる。

40

【0124】

図15にその例を示す。この図15では、N1(ナレーション1)とP1(家族)とが同時に提示されるべきという制約を有し、またN2(ナレーション2)とP2(息子)も同様な制約を有する。さらに、ビデオのタイムライン上に要素を配置するには、前述のナレーションとポーズとの関係のように、他の種類の制約も考慮する必要がある。

【0125】

セマンティクスを考慮したメディア同期は、コンテンツを構成する要素間の時間的制約(局所的制約)を記述したものであるといえ、時区間論理のモデルによって表現することが

50

できる。このモデルは、時間的関係の制約の記述形式としてよく知られたものであり、OWL-Sにおける時間に関するオントロジの一部にも組み込まれている。時区間論理のモデルを使えば、図15の要素間の制約は、以下のように記述される。このような制約は、たとえば、コンピュータ12の内部メモリ14(図1)に設定される。

【0126】

{VE1 overlaps P1、 VE1 meets N1、 N1 during P1、 N1 meets R1、 R1 meets VE2、 P1 overlaps VE2、 VE2 overlaps P2、 VE2 meets N2、 N2 meets R2、 N2 during P2}

{ 視覚効果または映像効果VE1は写真P1とオーバーラップする、映像効果VE1はナレーションN1の直前にある、ナレーションN1は写真P1の間にある、ナレーションN1は反応時間R1の直前にある、反応時間R1は映像効果VE2の直前にある、写真P1は映像効果VE2とオーバーラップする、映像効果VE2は写真P2とオーバーラップする、映像効果VE2はナレーションN2の直前にある、ナレーションN2は反応時間R2の直前にある、ナレーションN2は写真P2の間にある }

10

次に、上記の制約のもとで、思い出ビデオでこれらの各要素を提示する具体的なタイミングを計算する手続きを説明する。この計算のために、以下の条件を仮定する。なお、以下でサフィックスの*i*は、それが各タイプの*i*thの要素であることを示す。

条件1：写真トランジションにはフェードアウト/フェードインの視覚効果を用いる。また同じ写真内で人から人への遷移には、パン、ズームの視覚効果を適用する。

条件2：ビデオに提示される写真の順序はデフォルトで年代順とする。さらに各写真での人と人との間のパン、ズームについて、デフォルトで左から右とする。しかし製作者は必要に応じて、それを変更できる。

20

条件3：フェードアウト/フェードインに要する時間は任意に変更可能であるが、速過ぎる、または遅過ぎるフェードアウト/フェードインは、視聴者を不愉快にする可能性がある。そのために最小、最大の時間を予め与えておくことにする。以降、その最小時間をminTfoi_{*i*}とし、最大時間をmaxTfoi_{*i*}とする。

条件4：Tn_{*i*}はNarration_{*i*}に費やされる時間を表す。そして視聴者の予想される応答または反応に必要な時間を、Tr_{*i*}秒で表す。

条件5：パン(画素/秒)とズーム(倍/秒)の速度を変えることで、パン、ズームそれぞれに必要な時間を制御することができる。しかし、この効果についても、条件3と同様な理由から、最小、最大の時間を設定しておく必要がある。最小時間をminTpz_{*i*}とし、最大時間をmaxTpz_{*i*}として表す。

30

条件6：製作者は、思い出ビデオの最小、最大のビデオ再生時間を設定することができる。これらをminTpbとmaxTpbとしてそれぞれ表す。たとえば、製作者は、ビデオの長さを少なくとも30分以上で、高々60分以内のものになりたい等を指定することができる。

【0127】

このような条件のもとで、各要素のタイムライン上の並びを計算する。

【0128】

先に図2を参照して説明したように、思い出ビデオに使用する写真、および各写真の人々を表示する順番は、ステップS2において決定されている。したがって、それらに対する映像効果の要素は自動的に並べることができる。また、ステップS3での写真中の各人へのナレーションの選択は、それぞれの映像効果の要素に引き続いて実行される。この実施例において設定しているセマンティック同期のための制約(各要素間の局所的制約)では、そのナレーションに引き続いて、視聴者の応答または反応のためのポーズ時間を配置する。そして、視聴者の応答または反応があればそれに引き続いて、直ちに次の映像効果の要素が実行される。映像効果の要素は、提示順序が前後の写真中の、各人に対応する要素とオーバーラップする。同様の制約を適用して、映像効果要素、ナレーション要素および反応時間要素が最後の写真まで配置される。

40

【0129】

このようにして、図2のステップS2およびS3を繰り返し実行することによって、各要素のタイムラインをあらまし設計するが、図2のステップS5で実行するレンダリング

50

処理において、上で説明した同期の制約（条件 1 ないし 6）のもとで、コンピュータ 1 2 が、思い出ビデオのタイムライン上で各要素を実行する正確なタイミングを計算する。

【 0 1 3 0 】

まず、全ての要素（映像効果、ナレーション、反応）に対する、最小フェードアウト/フェードイン時間 $\min T_{foi_i}$ 、ナレーション時間 T_{n_i} 、反応時間 T_{r_i} 、および最小パンズーム時間 $\min T_{pz_i}$ の総和より最大ビデオ再生時間 $\max T_{pb}$ が小さいなら、すなわち、変数 i が 1 から n まで変化するときの $(\min T_{foi_i} + T_{n_i} + T_{r_i} + \min T_{pz_i}) > \max T_{pb}$ なら、図 2 のステップ S 5 において、コンピュータ 1 2 は、先に写真に付与しておいた優先順位に従って、優先順位の低い写真を自動的に除外する。ただし、製作者が手動で除外対象を選択するようにしてもよい。

10

【 0 1 3 1 】

また、全ての要素に対する、最大フェードアウト/フェードイン時間 $\max T_{foi_i}$ 、ナレーション時間 T_{n_i} 、反応時間 T_{r_i} 、および最大パンズーム時間 $\max T_{pz_i}$ の総和より最小ビデオ再生時間 $\min T_{pb}$ が大きいなら、つまり変数 i が 1 から n まで変化するときの $(\max T_{foi_i} + T_{n_i} + T_{r_i} + \max T_{pz_i}) < \min T_{pb}$ なら、ステップ S 5 のレンダリング処理において、コンピュータ 1 2 が幾つかの写真を追加するか、あるいは製作者が幾つかの写真を手動で追加する。追加する写真の画像データは図 2 のステップ S 1 において既に取り込まれているものを利用するか、あるいは必要に応じて追加的にその都度取り込むものとする。

【 0 1 3 2 】

そして、ビデオの再生時間長さが最大再生時間 $\max T_{pb}$ より短く、かつ最小再生時間 $\min T_{pb}$ より長くなるまで、写真の追加や削除による時間調整を行う。

20

【 0 1 3 3 】

このように、写真のアノテーションされた領域に対するケンバース効果などの映像効果と、オーディオの効果で特にナレーションとの間のセマンティックな同期制約を考慮することで、簡単に魅力的なスライドショービデオ（ビデオコンテンツ）を製作できる。具体的には、映像効果要素、ナレーション要素および反応時間要素を各写真に所定の条件でタイムライン上で割り付け、それら各要素の合計所要時間が最大ビデオ再生時間と最小ビデオ再生時間との間に入るように設計することによって、映像効果とナレーションとの同期および視聴者の反応時間が確実に確保できる。

【 図面の簡単な説明 】

30

【 0 1 3 4 】

【 図 1 】 この発明の一実施例のビデオコンテンツ作成装置を示すブロック図である。

【 図 2 】 図 1 実施例の全体動作を示すフロー図である。

【 図 3 】 図 1 実施例の G U I の一例を示す図解図である。

【 図 4 】 図 2 のステップ S 1 で利用する写真入力およびメタ情報登録時の G U I の表示の一例を示す図解図である。

【 図 5 】 図 4 においてリージョンの切り出しを示す図解図である。

【 図 6 】 図 5 のリージョンのメタ情報を示す図解図である。

【 図 7 】 写真のアノテーションを例示する図解図である。

【 図 8 】 ナレーションのアノテーションを例示する図解図である。

40

【 図 9 】 図 2 のステップ S 2 での写真の選択および再生順設定動作時の G U I の一例を示す図解図である。

【 図 1 0 】 図 2 のステップ S 5 での思い出ビデオのレンダリング動作を詳細に示すフロー図である。

【 図 1 1 】 図 8 の G U I におけるナレーション作成領域を示す図解図である。

【 図 1 2 】 図 2 のステップ S 7 でのナレーション付与動作を詳細に示すフロー図である。

【 図 1 3 】 ナレーション付与に用いるシソーラス辞書を示す図解図である。

【 図 1 4 】 ナレーション付与に用いる共起辞書を示す図解図である。

【 図 1 5 】 セマンティクスを考慮した要素のタイムライン上の配置の一例を示す図解図である。

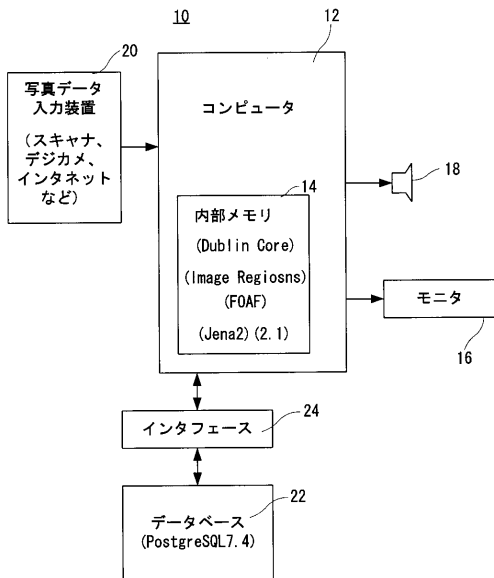
50

【符号の説明】

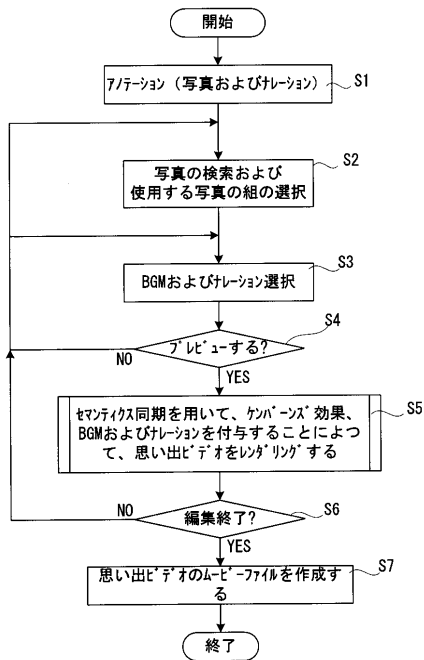
【0135】

- 10 ...ビデオコンテンツ作成装置
- 12 ...コンピュータ
- 14 ...内部メモリ
- 16 ...モニタ
- 20 ...写真データ入力装置
- 22 ...データベース
- 26 ...GUI
- 28 ...写真表示編集領域
- 32、34 ...メタ情報入力領域

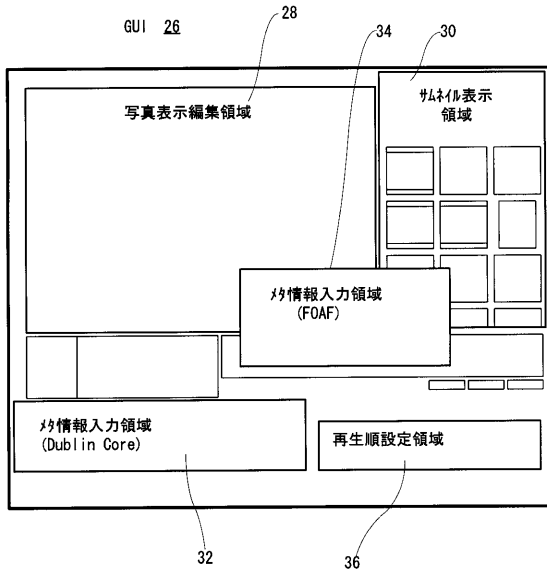
【図1】



【図2】



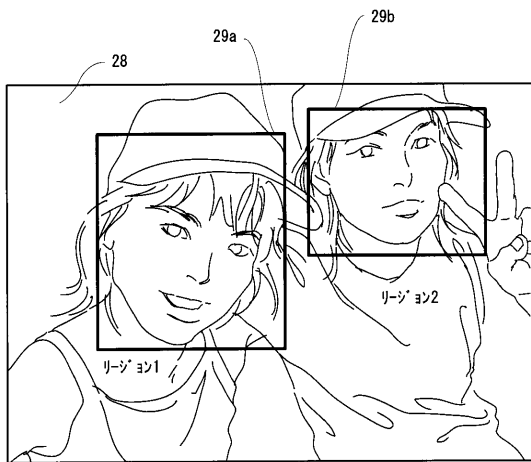
【図3】



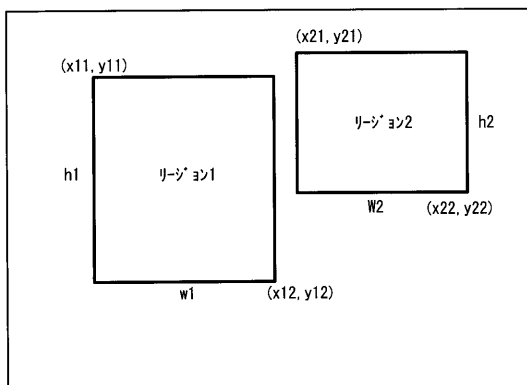
【図4】



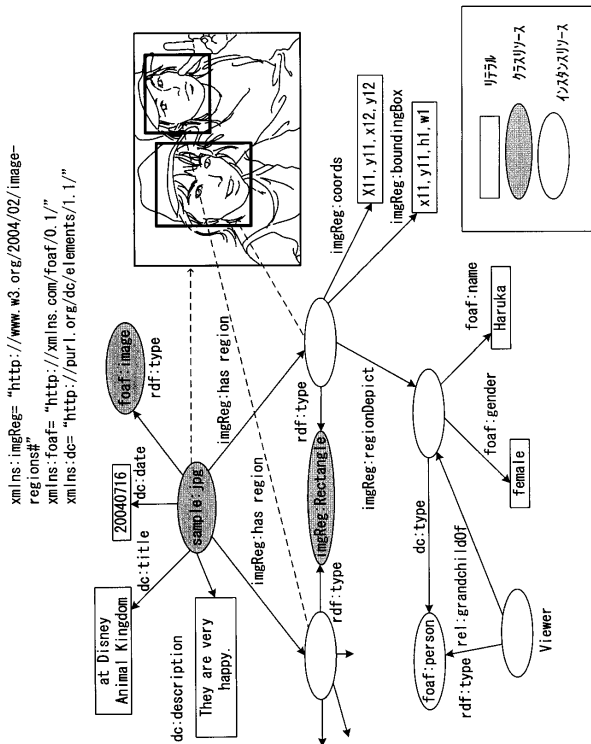
【図5】



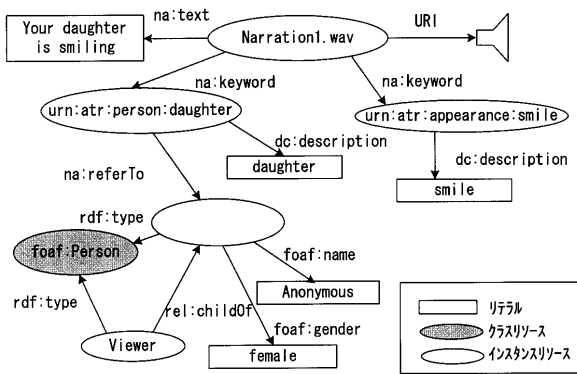
【図6】



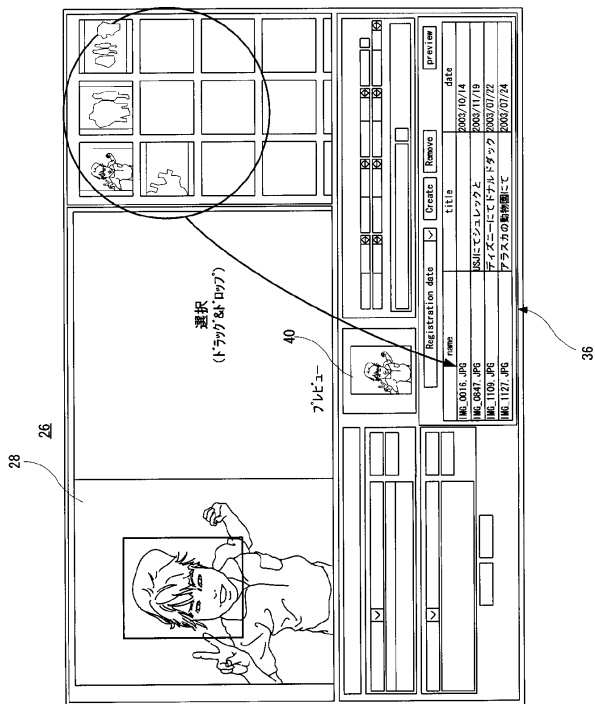
【図7】



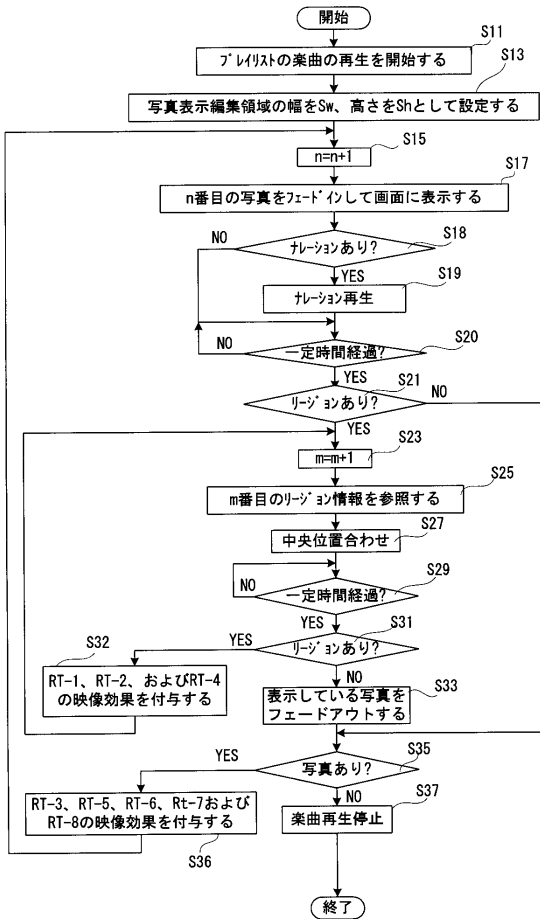
【図8】



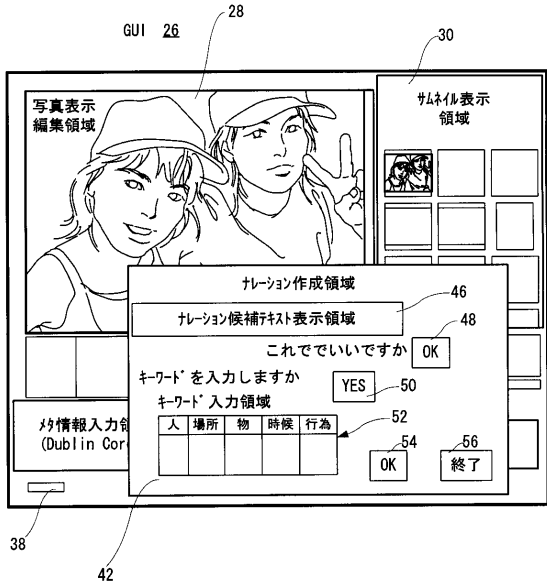
【図9】



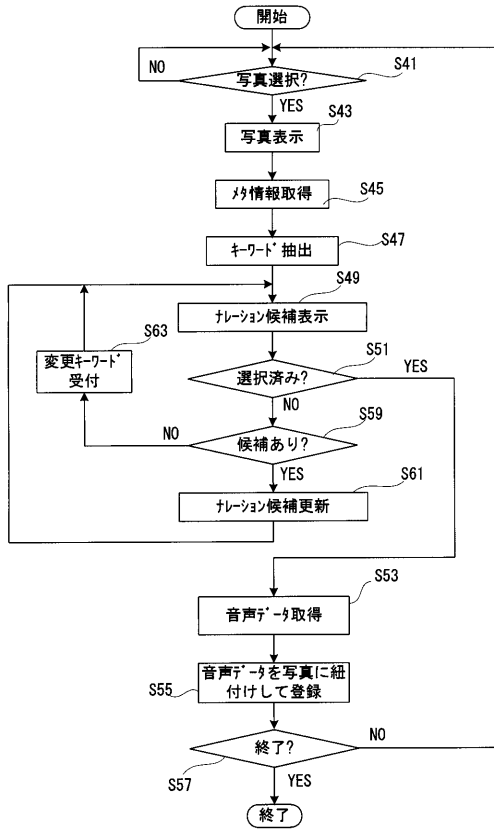
【図10】



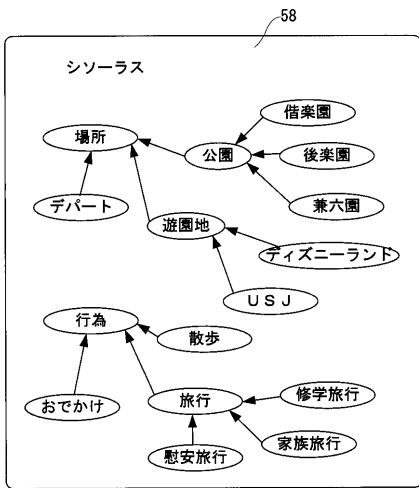
【図11】



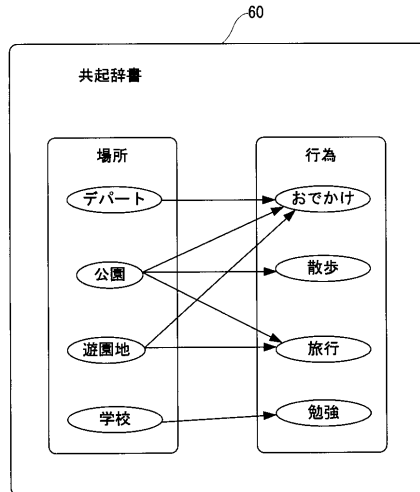
【図12】



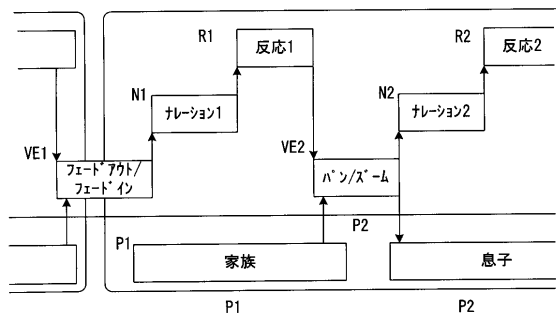
【図13】



【図14】



【図15】



フロントページの続き

特許権者において、実施許諾の用意がある。

(72)発明者 安田 清

京都府相楽郡精華町光台二丁目2番地2 株式会社国際電気通信基礎技術研究所内

審査官 竹中 辰利

(56)参考文献 特開2004-228731(JP,A)

特開2005-25715(JP,A)

特開2004-128570(JP,A)

(58)調査した分野(Int.Cl., DB名)

H04N 5/76 - 5/956

G06T 13/00