

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第5089996号
(P5089996)

(45) 発行日 平成24年12月5日 (2012. 12. 5)

(24) 登録日 平成24年9月21日 (2012. 9. 21)

(51) Int. Cl.	F 1
G 0 6 F 3/14 (2006.01)	G 0 6 F 3/14 3 1 0 Z
G 0 6 F 3/048 (2006.01)	G 0 6 F 3/048 6 5 0
G 0 6 F 3/16 (2006.01)	G 0 6 F 3/16 3 3 0 C

請求項の数 22 (全 42 頁)

(21) 出願番号	特願2007-5206 (P2007-5206)	(73) 特許権者	000006747
(22) 出願日	平成19年1月12日 (2007. 1. 12)		株式会社リコー
(65) 公開番号	特開2007-220093 (P2007-220093A)		東京都大田区中馬込 1 丁目 3 番 6 号
(43) 公開日	平成19年8月30日 (2007. 8. 30)	(74) 代理人	100070150
審査請求日	平成21年11月16日 (2009. 11. 16)		弁理士 伊東 忠彦
(31) 優先権主張番号	11/332, 533	(74) 代理人	100091214
(32) 優先日	平成18年1月13日 (2006. 1. 13)		弁理士 大貫 進介
(33) 優先権主張国	米国 (US)	(74) 代理人	100107766
			弁理士 伊東 忠重
		(72) 発明者	ベルナ エロール
			アメリカ合衆国, カリフォルニア 940
			25, メンロ・パーク, サンド・ヒル・ロ
			ード 2882番, スイート 115 リ
			コー イノベーション インク内

最終頁に続く

(54) 【発明の名称】 ナビゲーションパスの計算方法

(57) 【特許請求の範囲】

【請求項 1】

オーディブル、ビジュアル、及びオーディオビジュアル文書要素を受け取るステップと、

ディスプレイの制約に基づき、前記オーディブル、ビジュアル、及びオーディオビジュアル文書要素と関連する時間属性を計算するステップと、

前記時間属性と時間制約とに基づいて、マルチメディア表示に含めるオーディブル、ビジュアル、及びオーディオビジュアル文書要素のセットを選択するステップと、を有し、前記オーディオビジュアル文書要素は、図表部分を有するオーディオビジュアル図表要素を有し、前記時間属性を計算するステップは、前記オーディオビジュアル図表要素の時間属性を、前記図表部分をビジュアルに理解するのに要する時間と、前記オーディオビジュアル文書要素に付随する合成テキストを再生するのに要する時間との大きい方として計算するステップを含む、

ことを特徴とする方法。

【請求項 2】

1つのビジュアル文書要素はビジュアルテキスト要素を有し、時間属性の計算はそのビジュアルテキスト要素の時間属性の計算を有し、その計算は、ズームパラメータを計算するステップと、必要なパン動作を決定するステップと、

パンとズーム動作を実行するのに要する時間を決定するステップと、を有することを特徴とする、請求項 1 に記載の方法。

【請求項 3】

ビジュアルテキスト要素の時間属性の計算は、さらに、

ディスプレイ制約と文書情報に基づき第 1 のビューのビジュアルテキスト要素の可読性を決定するステップと、

ビジュアルテキスト要素が可読であると判断された場合、音声合成定数とビジュアルテキスト要素中の文字数の積に時間属性を設定するステップと、を有することを特徴とする、請求項 2 に記載の方法。

【請求項 4】

第 1 のビューで可読でない場合、

ズームビューでのビジュアルテキスト要素の可読性を決定するステップと、

ズームビューで可読であれば、音声合成定数とビジュアルテキスト要素中の文字数の積と、ズームビュー中のビジュアルテキスト要素中のタイトルにズームするのに必要な時間を示す定数とを足し合わせた値に、時間制約を設定するステップと、を有することを特徴とする、請求項 3 に記載の方法。

【請求項 5】

第 1 のビューまたはズームビューで可読でない場合、

ビジュアルテキスト要素中のテキストの可読性に必要なズーム量を示すズームファクタを計算するステップと、

パンに必要な時間を決定するステップと、

ズームビューのビジュアルテキスト要素のテキストにズームし、テキスト要素をパンするのに必要な時間に基づき時間属性を設定するステップと、を有することを特徴とする、請求項 4 に記載の方法。

【請求項 6】

ディスプレイ制約はディスプレイの幅と高さを含み、文書情報はページの幅と高さに基づくことを特徴とする、請求項 3 に記載の方法。

【請求項 7】

オーディオビジュアル文書要素は、テキスト部分を有するオーディオビジュアルテキスト要素を有し、

時間属性の計算は、オーディオビジュアル要素の時間属性を、オーディオビジュアル要素のテキスト部分をビジュアルにナビゲーションするのに要する時間と、そのオーディオビジュアル要素に付随する合成テキストを再生するのに要する時間との大きい方として、計算することを特徴とする、請求項 1 に記載の方法。

【請求項 8】

前記図表部分をビジュアルに理解するのに要する時間を、前記図表部分の図表タイプと、前記図表部分のエントロピーとに基づき計算する、請求項 1 に記載の方法。

【請求項 9】

前記オーディブル、ビジュアル、及びオーディオビジュアル文書要素のセットを用いて前記マルチメディア表示を作成するステップをさらに有することを特徴とする、請求項 1 に記載の方法。

【請求項 10】

オーディブル、ビジュアル、及びオーディオビジュアル文書要素のセットを用いてマルチメディア表示を作成するステップは、マルチメディア表示を生成するときオーディブル、ビジュアル、及びオーディオビジュアル文書要素のセットが処理されるナビゲーションパス表示を出力するステップを有することを特徴とする、請求項 9 に記載の方法。

【請求項 11】

時間属性と時間制約とに基づいて、マルチメディア表示に含めるオーディブル、ビジュアル、及びオーディオビジュアル文書要素のセットを選択するステップは、

ビジュアルとオーディオビジュアル文書要素の組み合わせか、オーディブルとオーディ

10

20

30

40

50

オブジュアル文書要素の組み合わせのいずれかを選択するステップと、

所与の時間制約の下で、含まれる情報を最大化するように文書要素の第1のセットを選択するステップと、

選択された文書要素を最大化した後、マルチメディア表示の残り時間の下で、含まれる情報を最大化するように文書要素の第2のセットを選択するステップと、を有することを特徴とする、請求項1に記載の方法。

【請求項12】

オーディブル、ビジュアル、及びオーディオビジュアル文書要素を受け取る手段と、
ディスプレイの制約に基づき、前記オーディブル、ビジュアル、及びオーディオビジュアル文書要素と関連する時間属性を計算する手段と、

前記時間及び情報コンテンツ属性と時間制約とに基づいて、マルチメディア表示に含めるオーディブル、ビジュアル、及びオーディオビジュアル文書要素のセットを選択する手段と、を有し、

前記オーディオビジュアル文書要素は、図表部分を有するオーディオビジュアル図表要素を有し、

前記時間属性を計算する手段は、前記オーディオビジュアル図表要素の時間属性を、前記図表部分をビジュアルに理解するのに要する時間と、前記オーディオビジュアル図表要素に付随する合成テキストを再生するのに要する時間との大きい方として計算する手段を含む、

ことを特徴とする情報処理装置。

【請求項13】

1つのビジュアル文書要素はビジュアルテキスト要素を有し、時間属性の計算はそのビジュアルテキスト要素の時間属性の計算を有し、その計算は、

ズームパラメータを計算する手段と、

必要なパン動作を決定する手段と、

パンとズーム動作を実行するのに要する時間を決定する手段と、を有することを、請求項12に記載の特徴とする情報処理装置。

【請求項14】

ビジュアルテキスト要素の時間属性の計算は、さらに、

ディスプレイ制約と文書情報に基づき第1のビューのビジュアルテキスト要素の可読性を決定する手段と、

ビジュアルテキスト要素が可読であると判断された場合、音声合成定数とビジュアルテキスト要素中の文字数の積に時間属性を設定する手段と、を有することを特徴とする、請求項13に記載の情報処理装置。

【請求項15】

第1のビューで可読でない場合、

ズームビューでのビジュアルテキスト要素の可読性を決定する手段と、

ズームビューで可読であれば、音声合成定数とビジュアルテキスト要素中の文字数の積と、ズームビュー中のビジュアルテキスト要素中のタイトルにズームするのに必要な時間を示す定数とを足し合わせた値に、時間制約を設定する手段と、を有することを特徴とする、請求項14に記載の情報処理装置。

【請求項16】

第1のビューまたはズームビューで可読でない場合、

ビジュアルテキスト要素中のテキストの可読性に必要なズーム量を示すズームファクタを計算する手段と、

パンに必要な時間を決定する手段と、

ズームビューのビジュアルテキスト要素のテキストにズームし、テキスト要素をパンするのに必要な時間に基づき時間属性を設定する手段と、を有することを特徴とする、請求項15に記載の情報処理装置。

【請求項17】

オーディオビジュアル文書要素は、テキスト部分を有するオーディオビジュアルテキスト要素を有し、

時間属性の計算は、オーディオビジュアル要素の時間属性を、オーディオビジュアル要素のテキスト部分をビジュアルにナビゲーションするのに要する時間と、そのオーディオビジュアル要素に付随する合成テキストを再生するのに要する時間との大きい方として、計算することを特徴とする、請求項 1 2 に記載の情報処理装置。

【請求項 1 8】

前記図表部分をビジュアルに理解するのに要する時間を、前記図表部分の図表タイプと、前記図表部分のエントロピーとに基づき計算する、請求項 1 2 に記載の情報処理装置。

【請求項 1 9】

前記オーディブル、ビジュアル、及びオーディオビジュアル文書要素のセットを用いて前記マルチメディア表示を作成する手段をさらに有することを特徴とする、請求項 1 2 に記載の情報処理装置。

【請求項 2 0】

オーディブル、ビジュアル、及びオーディオビジュアル文書要素のセットを用いてマルチメディア表示を作成する手段は、マルチメディア表示を生成するときにオーディブル、ビジュアル、及びオーディオビジュアル文書要素のセットが処理されるナビゲーションパス表示を出力する手段を有することを特徴とする、請求項 1 9 に記載の情報処理装置。

【請求項 2 1】

時間属性と時間制約とに基づいて、マルチメディア表示に含めるオーディブル、ビジュアル、及びオーディオビジュアル文書要素のセットを選択する手段は、

ビジュアルとオーディオビジュアル文書要素の組み合わせか、オーディブルとオーディオビジュアル文書要素の組み合わせのいずれかを選択する手段と、

所与の時間制約の下で、含まれる情報を最大化するように文書要素の第 1 のセットを選択する手段と、

選択された文書要素を最大化した後、マルチメディア表示の残り時間の下で、含まれる情報を最大化するように文書要素の第 2 のセットを選択する手段と、を有することを特徴とする、請求項 1 2 に記載の情報処理装置。

【請求項 2 2】

コンピュータに、請求項 1 ないし 1 1 いずれか一項記載の方法ステップを実行させることを特徴とするコンピュータプログラム。

【発明の詳細な説明】

【技術分野】

【0 0 0 1】

本発明は文書の処理と表示に関する。より具体的には、文書中に特定された可聴 (audible) 及び／または可視 (visual) 情報を有し、その文書の一部を表示する時に可聴情報を再生 (synthesized) する文書の作成に関する。

【背景技術】

【0 0 0 2】

本特許文献の開示には、その一部として (著作権または回路配置権の) 保護の対象となる素材が含まれている。(著作権または回路配置権の) 権利者は、特許庁の包帯または記録にある範囲において、本特許文献を複製することは認めるが、(著作権または回路配置権である) すべての権利を保持している。

【0 0 0 3】

ワイヤレスネットワーク、モバイルネットワーク、パーソナルモバイル機器等がどこでも使用できるようになりつつあり、多くの人々が小さなディスプレイと限定された入力装置を用いてウェブページ、写真、文書等をブラウズ (閲覧) している。現在のところ、小さなディスプレイを用いてウェブページを見るために、ウェブページを単純なグラフィックスのレベルが低いものにしている。写真を見るときにも、その解像度を低くしたものを表

10

20

30

40

50

示して、必要に応じて写真の一部を拡大したりスクロールしたりできるようにして、問題を解決している。

【0004】

一方、文書のブラウズに伴う問題はより困難である。文書は、ページが複数あり、写真よりも解像度が非常に高く（内容を見るためにユーザは拡大やスクロールをより頻繁に行わなければならない）、情報が非常に分散している（例えば、写真の焦点はフォーカスされている人の顔や被写体のみであるが、典型的な文書は多数の焦点を有する。例えば、タイトル、著者、要約、図面、参考文献等である。）。文書を見てブラウズする時の問題は、デスクトップやラップトップのディスプレイでは文書ビューアとブラウザを使用することにより解決されている。例えば、アドビアクロバット（www.adobe.com）やマイクロソフトワード（www.microsoft.com）等である。これらにより、文書の拡大（zooming）、文書ページ間の切り換え、サムネイルのスクロール等が可能となる。デスクトップアプリケーションではこのような非常にインタラクティブな処理が可能であるが、モバイル機器（例えば電話やPDAなど）は入力装置が限られており、ディスプレイも小さいので、これらの機器上で文書のブラウズ等をするためのよりよいソリューションが必要とされている。

10

【0005】

米国カリフォルニア州メンローパークにあるリコーイノベーション社は、ここでスマートネイルテクノロジーと呼ぶ技術を開発した。スマートネイルテクノロジーは、ディスプレイサイズの制限に合わせた別の画像表現を生成するものである。スマートネイル処理は、3つのステップを含む：（1）イメージセグメントを特定して、特定したイメージセグメントに解像度と重要性を付与するイメージ分析ステップと、（2）出力サムネイル中の可視コンテンツを選択するレイアウト決定ステップと、（3）選択されたイメージセグメントのクロッピング（cropping）、スケーリング、ペースティング（pasting）により最終的なスマートネイルイメージを作る作成ステップ。スマートネイル処理の入力は、出力と同様に、静止イメージである。すべての情報は、上記の3つのステップで処理されて、静的可視情報が得られる。より詳細な情報は、特許文献1、2に記載されている。

20

【0006】

ウェブページ要約（summarization）は、一般的に、ウェブページの要約を提供する先行技術として周知である。しかし、ウェブページ要約をする方法は、ほとんどがテキストに焦点を絞ったものであり、元のウェブページで使用されていない新しいチャンネル（例えばオーディオ）を組み込むものではない。例外として、目が不自由な人たちのためにオーディオを使用するものを以下に説明するが、特許文献3にも説明されている。

30

【0007】

Maderlechner等は、ユーザに対して、余白や文字の高さ等の文書の重要な特徴を調査し、自動的に文書の注意度の高い領域をセグメント化する注意ベースの文書モデルを開発した。これらの領域をハイライトし（例えば、これらの領域を暗くして、その他の領域をより透明にして）、ユーザが文書をより効率的にブラウズできるようにする。より詳細は、非特許文献1を参照せよ。

【0008】

先行技術の方法の一つは、モバイル機器上で非インタラクティブに画像をブラウズするものである。この方法は、画像上の顕著な領域、顔領域、テキスト領域を自動的に見つけて、ズーム及びパンをして見るものに自動的にクローズアップを見せるものである。この方法は、写真等の画像の再生にフォーカスしており、文書画像にはフォーカスしていない。よって、この方法は画像ベースのみであり、サムネイルのオーディオチャンネルを扱うものではない。また、我々のように各ビジュアルセグメントに関連した時間属性を有するものではない。より詳細は、非特許文献2及び3を参照せよ。

40

【0009】

先行技術による文書のオーディオへの変換は、目の不自由な人達を補助することにフォーカスしたものである。例えば、アドビ社は、PDF文書から音声を作成するアクロバット

50

トリーダーのプラグインを提供している。より詳細な情報は、目の不自由な方のためのPDFアクセス (<http://www.adobe.com/support/salesdocs/10446.htm>) を参照されたい。盲目または弱視の人のために文書からオーディオカセットの作り方に関するガイドラインがある。一般的なルールとして、表または写真の説明文に含まれる情報はオーディオカセットに含める。グラフィックスは一般的に省略する。より詳細は、「ヒューマンリソースツールボックス」モビリティインターナショナルUSA、2002年 (www.miusa.org/publications/Hrtoolboxintro.htm) を参照せよ。盲目及び弱視のユーザのためのブラウザを開発する仕事もされている。ある方法では、グラフィカルなHTML文書を3次元仮想サウンド空間環境にマッピングし、非音声の聴覚回復 (auditory cures) によりHTML文書を区別させる。より詳細は、非特許文献4を参照せよ。盲目または弱視のユーザのためのアプリケーションでは、必ずしもチャンネルを制約せずに、視覚的チャンネルを完全にあきらめないで、できるだけ多くの情報をオーディオチャンネルに変換することが目標となっているようである。

10

【0010】

メッセージの変換に使用するその他の先行技術による方法には、2001年6月19日に発行された特許文献3 (発明の名称「Wireless Delivery of Message Using Combination of Text and Voice」) がある。ここに説明したように、ハンドヘルド機器でユーザがボイスメールを受信するために、ボイスメールメッセージをフォーマットされたオーディオボイスメールメッセージとフォーマットされたテキストメッセージとに変換する。テキストに変換されるメッセージ部分をハンドヘルド機器のスクリーンに入力し、

20

メッセージの残りはオーディオとして設定される。

【特許文献1】米国特許出願第10/354,811号 (公開公報第2004/0146199A1号)

【特許文献2】米国特許出願第10/435,300号 (公開公報第2004/0145593A1号)

【特許文献3】米国特許公報第6,249,808号

【非特許文献1】Maderlechner等著「Information Extraction from Document Images using Attention Based Layout Segmentation」DLIA会報、ページ216-219、1999年

【非特許文献2】Wang等著「MobiPicture-Browsing Pictures on Mobile Devices」ACM MM '03、バークレー、2003年11月

30

【非特許文献3】Fan等著「Visual Attention Based Image Browsing on Mobile Devices」ICME、第1巻、ページ53-56、バルチモア、メリーランド、2003年7月

【非特許文献4】Roth等著「Auditory browser for blind and visually impaired users」CHI '99、ピッツバーグ、ペンシルバニア、1999年5月

【発明の開示】

【発明が解決しようとする課題】

【0011】

文書を可視化する方法、装置、及び製品を提供することである。

40

【課題を解決するための手段】

【0012】

一実施形態において、該発明は、オーディブル、ビジュアル、及びオーディオビジュアル文書要素を受け取るステップと、ディスプレイとアプリケーションの制約に基づき、前記オーディブル、ビジュアル、及びオーディオビジュアル文書要素と関連する時間属性を計算するステップと、ディスプレイの制約に基づき、前記オーディブル、ビジュアル、及びオーディオビジュアル文書要素と関連する情報コンテンツ属性を計算するステップと、前記時間及び情報コンテンツ属性に基づいて、前記1つ以上のオーディブル、ビジュアル、及びオーディオビジュアル文書要素のセットを選択するステップと、前記1つ以上のオーディブル、ビジュアル、及びオーディオビジュアル文書要素のセットを用いて前記マル

50

チメディア表示を作成するステップと、を有する。

【発明を実施するための最良の形態】

【0013】

本発明は、以下の詳細な説明と本発明のいろいろな実施形態を示した添付図面から、よりよく理解できるであろう。しかし、これらの実施形態は、本発明を限定されるものと解してはならず、説明と理解を目的としたものと解すべきである。

【0014】

ここでマルチメディアサムネイル（MMネイル）と呼ぶ、文書のマルチメディア概観を生成する方法と装置を説明する。この技術は、オーディオチャンネルとビジュアルチャンネル、及び空間次元と時間次元を利用して、ディスプレイが小さい機器上に複数ページの文書を表示するものである。これは、文書中で自動的にガイド付きツアーをするようなものである。

10

【0015】

一実施形態では、MMネイルは、文書の最も重要なビジュアル要素と可聴要素（例えばキーワード）を含み、これらの要素を空間領域と時間次元の両方で提示する。MMネイルは、出力機器に起因する制約（例えば、ディスプレイのサイズや画像描画能力上の制約など）やアプリケーションに起因する制約（例えば、オーディオ再生の時間上の制約など）を考慮して、情報を分析、選択、合成することにより得られる。

【0016】

一実施形態では、マルチメディア概観生成プロセスには、静止画像文書からマルチメディア情報（ビジュアル、オーディブル、及びオーディオビジュアル）を抽出し、ビジュアル、オーディブル、オーディオビジュアル情報セグメントに情報価値と時間属性を割り当て、機器及びアプリケーション上の制約があるビジュアル及びオーディオチャンネルに対してマルチメディア情報を最適化する。このように、MMネイルは、両方のチャンネルを使用することができる（必ずしも使用しなければならないのではない）。機器またはユーザの好みに応じて、2つのチャンネルのうちの一方が他方より好ましいかも知れない。結果として、MMネイルは、機器のビジュアルチャンネルとオーディオチャンネルの両方を用いて、限られたディスプレイと限られた時間フレームで文書の概観を提供し、ユーザが必要とするインタラクションを減らすか、潜在的には最小化する。

20

【0017】

MMネイルの一態様は、時間及びディスプレイに制約がある場合に、最終的な表現に含まれる文書要素を選択することである。一実施形態では、それぞれの文書要素の時間及び情報／関連性属性をまず計算し、最適化器を使ってMMネイルに含めるのにもっともよい要素を選択する。最適化フレームワークにより、ビジュアル文書要素とオーディオビジュアル文書要素の時間属性の新しい計算尺度を測る。一実施形態では、これらの尺度に対して最適化スキームを使用する。この最適化スキームはナップザックアルゴリズムの修正版である。

30

【0018】

以下の説明では、詳細に記載して本発明をより詳しく説明する。しかし、言うまでもなく、本発明はこれらの詳細がなくても実施することができる。他の場合では、詳細事項ではなくブロック図に周知の構造と機器を示すが、これは本発明が不明瞭になることを避けるためである。

40

【0019】

以下の詳細な説明の一部は、コンピュータメモリ中のデータビットに対する操作のアルゴリズムと記号による表現により表されている。これらのアルゴリズムによる説明と表現は、データ処理技術の当業者が、自分の仕事内容を他の分野の人に最も効果的に伝える手段である。ここで、また一般的に、アルゴリズムとは、所望の結果に導く自己矛盾のないステップのシーケンスである。このステップは、物理量の物理的操作を要するステップである。通常、必ずしも必要ではないが、この物理量には、記憶し、伝達し、結合し、比較し、操作できる電氣的または磁氣的信号の形をとる。主に一般的な使用のために、これら

50

の信号をビット、値、要素、記号、文字、式、数字等で表すと便利な時がある。

【0020】

しかし、これらの用語や類似の用語は適当な物理量と関連しているべきであり、これらの物理量に付された便利なラベルに過ぎないことに留意すべきである。特に断らなければ、以下の説明から明らかなように、言うまでもなく、この明細書全体において、「処理」、「算出」、「計算」、「判断」、「表示」等の用語を用いた説明は、コンピュータシステム、類似の電子的計算機器の動作やプロセスであって、コンピュータシステムのレジスタやメモリ内の物理的（電子的）量として表されたデータを操作し、コンピュータシステムメモリやレジスタ、その他の情報記憶装置、伝送機器、表示機器内の物理量として同様に表された他のデータに変換するものの動作や処理を指す。

10

【0021】

本発明は、また、これらの動作を実行する装置にも関する。この装置は、必要な目的のために特に構成されたものでもよく、コンピュータ中に記憶されたコンピュータプログラムにより選択的に起動または再構成された汎用コンピュータを有していてもよい。かかるコンピュータプログラムは、コンピュータによる読み取りが可能な記憶媒体に記憶することができる。このような記憶媒体には、例えば、フロッピー（登録商標）ディスク、光ディスク、CD-ROM、光磁気ディスク等のディスク、読出専用メモリ（ROM）、ランダムアクセスメモリ（RAM）、EPROM、EEPROM、磁気カード、光カード、その他の電子的命令の記憶に適した媒体であってコンピュータのシステムバスに結合されたものが含まれるが、これらには限定されない。

20

【0022】

ここで説明するアルゴリズムとディスプレイは、特定のコンピュータその他の装置に本質的に関係するものではない。いろいろな汎用システムをここでの教示に従ったプログラムで用いることができるし、必要な方法ステップを実行することに特化した装置を構成しても便利である。これらのシステムに必要な構成を以下に示す。また、本発明は特定のプログラミング言語により記述されるものではない。言うまでもなく、いろいろなプログラミング言語を用いてここに説明する本発明の教示を実施できる。

【0023】

機械読み取り可能媒体には、機械により読み取り可能な形式で情報を記憶または送信するメカニズムであれば、どんなものも含まれる。例えば、機械読み取り可能媒体には、読出専用メモリ（ROM）、ランダムアクセスメモリ（RAM）、磁気ディスク記憶媒体、光記憶媒体、フラッシュメモリデバイス、伝搬信号（電氣的・光学的・音響的その他の形式を含み、例えば搬送波・赤外線信号・デジタル信号等である）等が含まれる。

30

【0024】

概要

下記の選択スキームは、文書のビジュアル、オーディブル、及びオーディオビジュアル要素を取り、時間及び情報コンテンツ（例えば重要性）属性、時間、ディスプレイ、アプリケーションによる制約に基づき、文書用の組み合わせ及びナビゲーションパスを選択するものである。そうすることにより、文書のマルチメディア表現を生成する。

【0025】

図1は、文書のマルチメディア表示を生成するプロセスの一実施形態を示すフロー図である。このプロセスは、ハードウェア（例えば回路、専用ロジック等）、（汎用コンピュータシステムまたは専用機上で実行される）ソフトウェア、またはこれらの組み合わせを含む処理ロジックにより実行される。

40

【0026】

図1を参照して、このプロセスの開始において、処理ロジックが文書のオーディブル、ビジュアル、及びオーディオビジュアル要素を処理する（処理ブロック101）。オーディブル、ビジュアル、及びオーディオビジュアル文書要素をどのように得るかを以下に説明する。

【0027】

50

オーディブル、ビジュアル、及びオーディオビジュアル文書要素を用いて、処理ロジックは、ディスプレイと少なくとも1つのアプリケーションとユーザごとの制約に基づいて、オーディブル、ビジュアル、及びオーディオビジュアル文書要素と関連する時間属性を計算する（処理ブロック102）。一実施形態では、ディスプレイの制約は、マルチメディア表示を生成する機器のディスプレイサイズである。

【0028】

処理ロジックは、ディスプレイと少なくとも1つのアプリケーションとユーザごとの制約とに基づいて、オーディブル、ビジュアル、及びオーディオビジュアル文書要素に関連する情報コンテンツ（例えば重要性）属性の計算もする（処理ブロック103）。

【0029】

時間及び情報コンテンツ属性を計算した後、処理ロジックは、マルチメディア表示に含めるオーディブル、ビジュアル、オーディオビジュアル文書要素を、ディスプレイの制約と垂憂なくとも1つのアプリケーションとユーザごとの制約だけでなく時間及び情報コンテンツ属性に基づいて選択する（処理ブロック104）。一実施形態では、処理ロジックは、第1段階でビジュアルとオーディオビジュアル情報の組み合わせ、第2段階でオーディブル情報を選択し、いずれの場合も、処理ロジックは、文書要素の第1のセットを選択して、所与の時間的制約の下で第1段階でそれに含まれる情報を最大化し、選択された文書要素を最大化した後、文書要素の第2のセットを選択して、マルチメディア表示の時間的長さの残っている時間量の下で第2段階でその第2のセットに含まれている情報を最大化する。一実施形態では、1つ以上のアプリケーション及び／またはユーザ制約に基づく入力に基づいて、上記の選択をする。一実施形態では、どの段階が第1でありどの段階が第2であるかは、アプリケーション及び／またはユーザ制約の入力に基づく。一実施形態では、ページサムネイル制約の下で選択された文書要素の最大化をする。

【0030】

オーディブル、ビジュアル、及びオーディオビジュアル要素を選択した後、処理ロジックは、1つ以上のオーディブル、ビジュアル、オーディオビジュアル文書要素のセットを用いてマルチメディア表示を生成する（処理ブロック105）。一実施形態では、処理ロジックは、マルチメディア表示を生成するときに1つ以上のオーディブル、ビジュアル、オーディオビジュアル要素を処理するナビゲーションパスを出力することにより、そのマルチメディア表示を作成する。ナビゲーションパスは、限られたディスプレイ面積中にある時間で、いかにオーディブル、ビジュアル、及びオーディオビジュアル要素を提示するかを決めるものである。かかる要素間の遷移も決定する。ナビゲーションパスには、開始時間、文書要素の場所と大きさ、要素のフォーカス時間、文書要素間の遷移タイプ（例えば、パン、ズーム、フェードイン等）、遷移時間等に関して、要素の順序付けすることが含まれる。これには、オーディブル、ビジュアル、及びオーディオビジュアル文書要素を読む順序で並べ替えることが含まれる。

【0031】

オーディブル、ビジュアル、及びオーディオビジュアル文書要素の取得

一実施形態では、オーディブル、ビジュアル、及びオーディオビジュアル文書要素を生成するか、アナライザ（図示せず）を用いて取得する。アナライザが文書を受け取り、メタデータも受け取ることができる。メタデータには、著者情報、作成日付、（例えば、テキストがメタデータであり、文書画像にオーバーレイされるPDFファイルフォーマットの場合の）テキスト、オーディブルストリームまたはビデオストリーム、URL、出版者名、出版日、出版地、アクセス情報、暗号化情報、画像及びスキャン分解能、MPEG-7記述子等が含まれる。アナライザは、これらの入力に応答して、これらの入力に前処理を実施し、文書中の1つ以上のビジュアルフォーカスポイントを示す出力情報と、文書中のオーディブル情報を示す情報と、文書中のオーディオビジュアル情報を示す情報とを生成する。文書要素から抽出した情報がビジュアル情報とオーディブル情報を示す場合、この要素はオーディオビジュアル要素の候補である。アプリケーションまたはユーザが、候補のセットから得たオーディオビジュアル要素を最終的に選択してもよい。オーディオビ

10

20

30

40

50

ジュアル要素中のオーディブル情報とビジュアル情報は同期していても（していなくても）よい。例えば、アプリケーションは、文書中の図表とその注釈が同期していることを要する。オーディブル情報は、文書及び／またはメタデータ中の重要な情報であってもよい。

【0032】

一実施形態では、アナライザは、文書前処理部、メタデータ前処理部、ビジュアルフォーカスポイント識別器、重要オーディブル文書情報識別器、及びオーディオビジュアル情報識別器を有する。一実施形態では、文書前処理部は、1つ以上の光学的文字認識（OCR）と、レイアウト分析と、レイアウト抽出と、JPEG2000圧縮と、ヘッダー抽出と、文書フロー分析と、フォント抽出と、顔検出と、顔認識と、グラフィックス抽出と、音符認識とのうち1つ以上を実行する。どれを実行するかはアプリケーションにより異なる。一実施形態では、文書前処理部は、ExpervisionOCRソフトウェア（詳細はwww.expervision.comから得られる）を含み、文字のレイアウト分析を行い、境界ボックスと、フォントサイズやフォントタイプ等の関連属性とを生成する。他の実施形態では、テキスト領域の囲みボックスと関連属性を、ScanSoftソフトウェア（詳細はwww.nuance.comから得られる）を用いて生成する。他の実施形態では、Aiello M., Monz, C., Todoran, L., Worring, M. 著、「Document Understanding for a Broad Class of Documents」（International Journal on Document Analysis and Recognition (IJ DAR), vol.5(1), pp. 1-16, 2002）に記載されたように、テキスト領域の意味分析を実施して、タイトル、ヘッダー、フッター、図面注釈等の意味属性を決定する。

10

20

【0033】

メタデータ前処理部は、構文解析とコンテンツ収集を実行する。例えば、一実施形態では、メタデータ前処理部は、著者名をメタデータとして与えられると、ワールドワイドウェブ（WWW）から著者の写真を抽出する（これは後でMMネイルに含めることができる）。一実施形態では、メタデータ前処理部はXML構文解析を実行する。

【0034】

前処理が済むと、ビジュアルフォーカスポイント識別器がビジュアルフォーカスセグメントを抽出し、一方、重要オーディブル文書情報識別器が重要オーディオデータを抽出し、オーディオビジュアル情報識別器が重要オーディオビジュアルデータを決定して抽出する。

30

【0035】

一実施形態では、ビジュアルフォーカスポイント識別器が、OCRと前処理部からのレイアウト分析結果と前処理部からのXML構文解析結果に基づいて、ビジュアルフォーカスポイントを識別する。

【0036】

一実施形態では、ビジュアルフォーカスポイント（VTP）識別器は、2003年5月9日出願の米国特許出願第10/435,300号（発明の名称「Resolution Sensitive Layout of Document Regions」、2004年7月29日公開、公開番号US2004/0145593A1号公報）に記載された分析手法を実行して、テキスト領域とそれに付随する属性（例えば、重要性及び解像度属性）を識別する。テキスト領域はタイトルと見だしを含む。タイトルや見だしはセグメントとして解釈される。一実施形態では、ビジュアルフォーカスポイント識別器は、タイトルと図表も判定する。一実施形態では、図表をセグメント化する。

40

【0037】

一実施形態では、オーディブル文書情報（ADI）識別器が、OCRと前処理部からのレイアウト分析結果と前処理部からのXML構文解析結果に応じて、オーディブル情報を識別する。

【0038】

ビジュアルフォーカスセグメントの例としては、図表、タイトル、フォントが大きいテキスト、人が写っている写真などがある。これらのビジュアルフォーカスポイントはアブ

50

リケーションによって異なることには留意されたい。また、解像度及び特徴属性等の属性もこのデータに関連する。解像度はメタデータとして指定されてもよい。一実施形態では、これらのビジュアルフォーカスセグメントは、2003年5月9日出願の米国特許出願第10/435,300号（発明の名称「Resolution Sensitive Layout of Document Regions」、2004年7月29日公開、公開番号第US2004/0145593A1号公報）に記載されたのと同じ方法で判定される。他の実施形態では、ビジュアルフォーカスセグメントは、Le Meur,O.,Le Callet,P.,Barba,D.,Thoreau,D.著「Performance assessment of a visual attention system entirely based on a human vision modeling」（ICIP 2004会報、シンガポール、ページ2327-2330、2004年）に記載されたのと同じ方法で判定される。特徴はビジュアルセグメントのタイプによって異なってもよい（例えば、フォントが大きいテキストはフォントが小さいテキストよりも重要であったり、アプリケーションによってはその逆であったりする）。これらのセグメントの重要性は、MMネイルを生成する前に、各アプリケーションについて経験的に決定される。例えば、経験的研究により、ユーザが文書のスキャン品質を評価するアプリケーションでは、画像中の顔や短いテキストが最も重要なビジュアルポイントであることが分かるかも知れない。特徴ポイントを見つけるには、先行技術の文書及び画像分析方法の1つを用いることもできる。

10

【0039】

オーディブル情報の例としては、タイトル、画像注釈、キーワード、構文解析されたメタデータが含まれる。情報コンテンツ、関連性（特徴）、時間属性（音声合成後の時間的長さ）等の属性もオーディブル情報に付加される。オーディブルセグメントの情報コンテンツはそのタイプにより異なる。例えば、経験的な研究により、「文書サマリーアプリケーション」の場合、文書タイトルと画像注釈が文書中の最も重要なオーディブル情報であることが分かるかも知れない。

20

【0040】

VFPとADIの属性をクロス分析を用いて割り当てられる。例えば、画像（VFP）の時間属性を割り当てて、画像注釈（ADI）の時間属性と同じとすることができる。

【0041】

一実施形態では、オーディブル文書情報識別器は、TFIDF分析を実行して頻度に基づいてキーワードを自動的に決定する。この方法は、例えば次の文献に記載されているMatsuo, Y., Ishizuka, M.著「Keyword Extraction from a Single Document using Word Co-occurrence Statistical Information」（International Journal on Artificial Intelligence Tools, vol.13, no.1, pp.157-169, 2004）、またはFukumoto,F.,Suzuki,Y., ukumoto,J.著「An Automatic Extraction of Key Paragraphs Based on Context Dependency」（Proceedings of Fifth Conference on Applied Natural Language Processing, pp.291-298, 1997）の主要段落に記載されている。各キーワードについて、オーディブル文書情報識別器は、時間属性を、シンセサイザーがそのキーワードを話すのにかかる時間として計算する。

30

【0042】

同様に、オーディブル文書情報識別器は、タイトル、ヘッディング（headings）、画像注釈等の選択されたテキスト領域の時間属性を計算する。各時間属性は対応するセグメントと関連している。例えば、画像注釈時間属性は、対応する画像セグメントと関連している。一実施形態では、各オーディブル情報セグメントは、（フォントサイズやページ中での位置に基づく）ビジュアルな重要性や、テキスト領域の場合には読む順序や、キーワードの場合には出現頻度や、画像のビジュアルな重要性属性や、関連する画像注釈を反映する情報コンテンツ属性も担っている。一実施形態では、これらの情報コンテンツ属性は、2003年5月9日出願の米国特許出願第10/435,300号（発明の名称「Resolution Sensitive Layout of Document Regions」、2004年7月29日公開、公開番号第US2004/0145593A1号公報）に記載されたのと同じ方法で計算される。

40

50

【0043】

オーディオビジュアル文書情報（A V D I）は、オーディオビジュアル要素から抽出する情報である。

【0044】

このように、一実施形態では、文書の電子版（必ずしもビデオやオーディオのデータは含まないもの）とそのメタデータを用いて、ビジュアルフォーカスポイント（V F P）、重要オーディブル文書情報（A D I）、オーディオビジュアル文書情報（A V D I）を決定してもよい。

【0045】

ビジュアルフォーカスセグメント、重要オーディブル情報、オーディオビジュアル情報を最適化器に与える。V F P、A D I、及びA V D Iを装置及びアプリケーション制約（ディスプレイサイズ、時間制約等）とともに与えると、最適化器は出力表示（例えばマルチメディアサムネイル）に含むべき情報を選択する。一実施形態では、好ましいビジュアル、オーディブル、及びオーディオビジュアル情報を出力表示に含めるように選択を最適化する。ここで、好ましい情報には、文書中の重要情報、ユーザが好む、重要なビジュアル情報（例えば画像）、重要な意味情報（例えばタイトル）、キーパラグラフ（意味分析の出力）、文書コンテキストを含んでもよい。重要情報には、文書の解像度敏感領域を含んでもよい。選択は、計算された時間属性と情報コンテンツ（例えば重要性）属性に基づく。

10

【0046】

図2は、ビジュアル情報、可聴情報、オーディオビジュアル情報を処理するプロセスの一実施形態を示すフロー図である。このプロセスは、ハードウェア（例えば回路、専用ロジック等）、（汎用コンピュータシステムまたは専用機上で実行される）ソフトウェア、またはこれらの組み合わせを含む処理ロジックにより実行される。

20

【0047】

図2を参照して、本プロセスの開始において、処理ロジックは、ディスプレイ制約210とアプリケーション／タスク及びユーザ毎の制約211を用いてオーディブル文書要素201、ビジュアル文書要素202、オーディオビジュアル文書要素203と関連する時間属性を計算する（処理ブロック204）。また、処理ロジックは、ディスプレイ制約210とアプリケーション／タスク制約211を用いてオーディブル文書要素201、ビジュアル文書要素202、オーディオビジュアル文書要素203と関連する情報コンテンツ（例えば重要性）属性を計算する（処理ブロック206）。最適化器206中の処理ロジックは、計算された時間属性、計算された情報コンテンツ属性、及び時間制約212を受け取り、これに応じて、ナビゲーションパス207を生成する。

30

【0048】

選択後、この情報から、（ビジュアルチャネルまたは表示用の）ビジュアルデータ、（オーディオチャネルまたは表示用の）オーディオデータ、（オーディオチャネル及びビジュアルチャネルの組み合わせまたは表示用の）オーディオビジュアルデータを合成しマルチメディアサムネイルを生成する。よってこの場合、サムネイルは、文書の素材をビジュアルかつオーディブルに提示する表示のことを言う。

40

【0049】

以下、これらの動作をより詳細に説明する。ビジュアル要素の時間属性の計算、オーディオビジュアル要素の時間属性の計算、ページ制約がある2段階最適化アルゴリズム等を説明する。

【0050】

時間属性の計算

マルチメディア表示の文書要素選択の最適化には、可読性のためのレイアウトとサイズの最適化や余白の縮小等の空間的制約が係わる。かかるフレームワークでは、いくつかの情報コンテンツ（意味的、ビジュアル）属性が共通に文書要素と関連する。ここで説明したフレームワークにおいて、一実施形態では、空間的表示と時間的表示の両方を最適化す

50

る。そのため、「時間属性」が文書要素に付随する。以下のセクションでは、オーディブル、ビジュアル、及びオーディオビジュアル文書要素の時間属性の割り当てを詳細に説明する。

【0051】

オーディブル要素の時間属性 $t_{a,j}$ の計算

文書のオーディブル要素のフォーマットはいろいろあり、テキスト、波形、音符等がある。一実施形態では、波形または音符を再生するための最適時間は一般的にその表示に固有であり（例えば、波形の時間軸等）、これらの要素にその時間属性として自動的に付随することができる。

【0052】

テキストフォームのオーディブル要素の場合、一実施形態では、時間属性は次式のように計算する：

$$t_{a,j} = SSC \times N \quad (\text{式1})$$

ここで、Nはテキスト中の文字（数字、符号を含んでもよい）数であり、SSC（音声合成定数）は合成音声の各文字の再生に要する平均時間である。一実施形態では、SSCは次のように計算される：（1）文字数Kが分かっているテキスト文書を合成し、（2）合成した音声の発話に係るトータル時間Tを測定し、 $SSC = T / K$ とする。SSC定数は、言語、使用するシンセサイザー、及びそのオプション（女声と男声、アクセントのタイプ、話す速さ等）に応じて変化してもよい。一実施形態では、AT&TスピーチSDK（<http://www.nextup.com/attnv.html>から入手できる）を用いると、女声の英語の場合SSCは75msと計算された。

【0053】

ビジュアル要素の時間属性 $t_{v,j}$ の計算

文書のビジュアル要素を、「テキスト」要素と、画像、表、グラフ等の「画像」要素等との2つのセットにグループ分けすることができる。

【0054】

一実施形態では、テキストビジュアル要素の時間属性は、可読な解像度でユーザにテキストセグメントを見せるのに必要なビジュアル効果の時間的長さである。

【0055】

図3は、文書のビジュアルテキスト文書要素の時間属性を計算するプロセスの一実施形態を示すフロー図である。このプロセスは、ハードウェア（例えば回路、専用ロジック等）、（汎用コンピュータシステムまたは専用機上で実行される）ソフトウェア、またはこれらの組み合わせを含む処理ロジックにより実行される。

【0056】

図3を参照して、このプロセスの開始において、処理ロジックが文書のビジュアルテキスト文書要素を受け取る（処理ブロック301）。ビジュアルテキスト文書要素を用いて、処理ロジックはビジュアルテキスト要素の時間属性を計算する（処理ブロック302）。この計算は、ビジュアルテキスト要素の可読性解像度を計算し（処理ブロック302A）、ビジュアルテキスト要素を読むために必要なパン及びズーム動作を決定し（処理ブロック302B）、可読性解像度でその動作を実行するために必要な時間を決定する（処理ブロック302C）。

【0057】

このアルゴリズムの一実施形態の一般的概要を図4に示し、擬似コードの例を下に示す：

w_d ：表示幅（DC）

h_d ：表示高さ（DC）

w_p ：ページ幅（DI）

h_p ：ページ高さ（DI）

$w_{v,i}$

$h_{v,i}$ ：1テキストラインの元の幅と高さ（DEI）

10

20

30

40

50

N：テキスト中の文字数（D E I）

MRH：最小可読性高さ（＝10ピクセル）（D C）

ZOOMC：：タイトルズーム定数（＝0.5sec）（A C）

PANC：：パンピクセル定数（＝0.01sec）（A C）

【0058】

【表 1】

```

 $t_{v,j}=0, h_{v,i} = h_{v,i}$ 

if ( $\frac{w_d h_{v,i}}{w_p} \geq MRH$ ) //サムネイルビューでテキストが見えるか最初にテストする。テキス
    トが見えれば、ズームインする必要はない
{
 $t_{v,j} = SCC \times N$ 
}

else
{
 $h'_{v,i} = \frac{w_d h_{v,i}}{w_{v,i}}$  //テキスト幅がディスプレイ幅を満たすように新しい高さを計算する。
}

if ( $h_{v,i} \geq MRH$ ) //新しい高さが十分高いかテストする
{
 $t_{v,j} = ZOOMC + SCC \times N$  //テキストにズームして保持する
}

else
{
 $w'_{v,i} = \frac{w_{v,i} MRH}{h_{v,i}}$  //新しい幅を計算する
 $diffw = w'_{v,i} - w_d$  //ディスプレイにフィットしない残りの幅(ディスプレ
イに納まりきらない幅)を計算する
 $initN = \frac{diffw}{w'_{v,i}} \times N$ 
 $t_{v,j} = ZOOMC + SCC \times N' + \max(diffw \times PANC, SCC \times N'')$ 
 $N' \leq N - initN, N'' \leq initN$ 
}
}

END

```

上記のコードは、単一のテキストラインがあることを仮定している。このコードは、すべてのテキストラインの高さ $ht_{v,i}$ を考慮することによって、複数ラインの場合に適合させられる。また、上記のコードでは、DCはディスプレイ制約、DIは文書情報、DEIは文書要素情報、ACはアプリケーション制約であることに注意する。

【0059】

図4は、文書のビジュアルテキスト文書要素の時間属性を計算するプロセスの他の実施形態を示すフロー図である。このプロセスは、ハードウェア（例えば回路、専用ロジック等）、（汎用コンピュータシステムまたは専用機上で実行される）ソフトウェア、またはこれらの組み合わせを含む処理ロジックにより実行される。

【0060】

図4を参照して、このプロセスの開始において、処理ロジックが文書のビジュアルテキスト文書要素を受け取る（処理ブロック401）。処理ロジックは、最初のビューにおいてビジュアルテキスト文書要素が読み取れるかを、ディスプレイ制約、アプリケーション制約、及び文書情報に基づいて判断する。

10

【0061】

一実施形態では、最初のビューはサムネイルビューを有する。一実施形態では、ディスプレイ制約はディスプレイの幅と高さを含み、文書情報はページの幅と高さに基づく。この場合には、処理ロジックは、ビジュアルテキスト要素が可読であると判断されると、時間制約を音声合成定数とビジュアルテキスト要素中の文字数の積に設定する（処理ロジック403）。

【0062】

最初のビューで可読でなければ、処理ロジックはビジュアルテキスト要素がズームビューで可読か判断する（処理ロジック404）。その場合、処理ロジックは、ズームビューで可読であれば、音声合成定数とビジュアルテキスト要素中の文字数の積とズームビュー中のビジュアルテキスト要素中のタイトルにズームするのに必要な時間を示す制約とを足し合わせた値に、時間制約を設定する。

20

【0063】

ビジュアルテキスト要素が最初のビューまたはズームビューで可読でなければ、処理ロジックは、ビジュアルテキスト要素中のテキストの可読性に必要なズーム量を示すズームファクタを計算し（処理ロジック406）、パンに要する時間を判断する。そして、時間属性を次の値（第1ないし第3の値）の和に設定する（処理ロジック407）。すなわち、パンピクセル定数と、1ラインの元のテキストの幅・高さ及びディスプレイの幅・高さの間の差との積（第1の値）、ズームビュー中のビジュアルテキスト要素中のタイトルにズームするのに必要な時間を示す定数（第2の値）、音声合成定数と、最小可読高さに対するディスプレイ高さの商をビジュアルテキスト要素中の文字数から引いた結果との積（第3の値）、の和とする。

30

【0064】

他の実施形態では、処理ロジックは、最小可読高さ（MRH）にズームして、要素の幅がディスプレイの幅にフィットするかチェックする。フィットすれば、それを受け入れる。フィットしなければ、パン（panning）をする。ズーム定数を固定する必要はない。例えば、ズーム定数は次式で計算してもよい：

【0065】

【数1】

40

$$ZOOMC = \frac{MRH}{h_{v,i}} \times (a \text{ zoom speed factor}) ,$$

さらに他の実施形態では、ビジュアルテキスト要素eの時間属性を計算するため、最初に文書イメージをダウンサンプリングしてディスプレイ領域にフィットさせる。次に、Z(

50

e)をテキスト中の最小フォントの高さを最小可読高さにするのに必要なズームファクタとして決定する。最後に、ビジュアルテキスト要素eの時間属性を次のように計算する：

【0066】

【数2】

$$t(e) = \begin{cases} SCC \cdot n_e, & Z(e) = 1 \\ SCC \times n_e + Z_c, & Z(e) > 1 \end{cases},$$

10

ここで、 n_e はe（ビジュアルテキスト要素）中の文字数であり、 Z_c はズーム時間（我々の実施形態では1秒に固定している）であり、SSC（音声合成定数）は合成されたオーディオ（audio character）を再生するのに要する平均時間である。テキスト要素を1回のズーム動作で表示できず、ズームとパンの両方の動作が必要であっても、 $t(e)$ の計算は同じである。かかる場合、プレゼンテーション時間は、最初にテキストの一部（例えば最初のm文字）にズームし、そのテキスト上へのフォーカスを $SCC \times m$ 秒持続して使用される。次に、残り時間、すなわち $SCC \times (n_e - m)$ パン動作に使う。

【0067】

図表要素の場合、一実施形態では、ユーザがその文書要素を理解するのに要する予測時間に基づき時間制約を計算する。理解時間CTは、次式の通り、図表タイプと図表要素のエントロピー $e_{v,i}$ により決まる：

20

【0068】

【数3】

$$CT = FC(e_{vi}) \times E(e_{vi}),$$

(式2)

ここで、 $FC(e_{v,i})$ は図表定数であり、図表タイプ（画像、テーブル、グラフ等）により決まり、 $E(e_{v,i})$ は図表の規格化（正規化）エントロピー（normalized entropy）である。一実施形態では、規格化(normalization)は次のように実行される：

30

[図表のエントロピー] / [ページ全体のエントロピー]

特許文献1、すなわち米国特許出願第10／354,811号（公開公報第2004／0146199A1号）と、特許文献2、すなわち米国特許出願第10／435,300号（公開公報第2004／0145593A1号）を参照されたい。パラメータ $FC(e_{v,i})$ を使用する理由は、タイプが異なる図表は理解するのにかかる時間も異なるからである。例えば、一般的に、画像はグラフよりも理解しやすい。それゆえ、タイプが異なる図表の定数はそれぞれ異なる。グラフとテーブルの場合、図表を理解するのにかかる時間は、ビジュアル要素に含まれる文字数、単語数、及び数字の数により異なるかも知れない。また、CTはこれらの文書要素のビジュアルな複雑さにより異なる。この複雑さはその文書要素のエントロピーに基づく。

40

【0069】

オーディオビジュアル要素の時間属性 $e_{av,i}$ の計算

オーディオビジュアル要素は同期したオーディオ成分とビジュアル成分を含む文書要素である。例としては、注釈付きの図表であって、ビジュアル部分を有し、その図表注釈の合成音声オーディオ部分であるものがある。一実施形態では、オーディオビジュアル要素の時間属性は、オーディオビジュアルテキスト要素の場合と図表要素の場合とは異なる。以下説明する。

【0070】

50

テキストタイプのオーディオビジュアル要素は2つの成分を含む：

$V(e_{av,i})$ ： $e_{av,i}$ のビジュアル成分

$A(e_{av,i})$ ： $e_{av,i}$ のオーディオ成分であり、合成テキストである

テキストタイプのオーディオビジュアル要素の場合、一実施形態では、時間属性を次のように計算する：

【0071】

【数4】

$$t_{av,i} = \max\{t_{V(e_{av,i})}, t_{A(e_{av,i})}\} \quad (式3)$$

10

変数

【0072】

【数5】

$t_{V(e_{av,i})}$

20

は要素のビジュアル部分 $e_{av,i}$ の時間属性である。一実施形態では、上記のコードを用いてこれを計算する。変数

【0073】

【数6】

$t_{A(e_{av,i})}$

30

は要素のオーディブル部分 $e_{av,i}$ の時間属性である。一実施形態では、上記の式1としてSCCを用いてこれを計算する。

【0074】

図表タイプのオーディオビジュアル要素は2つの成分を含む：

$V(e_{av,i})$ ： $e_{av,i}$ のビジュアル成分であり、例えば図表、テーブル、グラフ等である。

【0075】

$A(e_{av,i})$ ： $e_{av,i}$ のオーディオ成分であり、例えば図表注釈や図表中のテキスト等である。

40

【0076】

しかし、図表注釈は長すぎる場合がほとんどである。注釈を短くする方法は幾つかある。例えば、注釈がいくつかの文から構成されている場合、最初の文のみを合成する。注釈が(a)、(b)等の多数のサブセクションを有するときは、これらのサブセクションが合成されるまでの部分のみを合成する。

【0077】

図表タイプのオーディオビジュアル要素の場合、一実施形態では、時間属性を次のように計算する：

【0078】

【数 7】

$$t_{av,i} = \max\{CT(V(e_{av,i})), t_{A(c_{av,i})}\} \quad (式 4)$$

$CT(V(e_{av,i}))$ は、上記の式 2 に示した、要素のビジュアル部分 $e_{av,i}$ の図表を理解する時間である。変数

【0079】

10

【数 8】

$$t_{A(c_{av,i})}$$

は要素のオーディブル部分 $e_{av,i}$ の時間属性である。一実施形態では、上記の式 1 に示したようにこれを計算する。

【0080】

図 5 は、文書のオーディオビジュアル文書要素の時間属性を生成するプロセスの一実施形態を示すフロー図である。このプロセスは、ハードウェア（例えば回路、専用ロジック等）、（汎用コンピュータシステムまたは専用機上で実行される）ソフトウェア、またはこれらの組み合わせを含む処理ロジックにより実行される。

20

【0081】

図 5 を参照して、このプロセスの開始において、処理ロジックが文書のオーディオビジュアル文書要素を受け取る（処理ブロック 501）。

【0082】

次に、処理ロジックは、オーディオビジュアル文書要素のビジュアル部分をビジュアルにナビゲーションするのに要する時間と、オーディオビジュアル文書要素の合成部分を再生するのに要する時間の大きい方である時間属性を計算する。

30

【0083】

一実施形態では、これの実行は、処理ロジックが、オーディオビジュアル文書要素がオーディオビジュアルテキスト要素か、それともオーディオビジュアル図表要素かを判断することにより行う。オーディオビジュアル文書要素がオーディオビジュアルテキスト要素である場合、処理ロジックは、オーディオビジュアル要素の時間属性を、オーディオビジュアル要素のテキスト部分をビジュアルにナビゲーションするのに要する時間と、そのオーディオビジュアル要素に付随する合成テキストを再生するのに要する時間との大きい方として、計算する。オーディオビジュアル文書要素がオーディオビジュアル図表要素である場合、処理ロジックは、オーディオビジュアル要素の時間属性を、オーディオビジュアル要素の図表部分をビジュアルに理解するのに要する時間と、そのオーディオビジュアル要素に付随する合成テキストを再生するのに要する時間との大きい方として、計算する。

40

【0084】

時間属性のまとめ

下の表 1 は、マルチメディアサムネイルに含めることができる異なる文書要素と、その要素タイプと、各要素タイプの場合に時間属性をいかに計算するかをまとめたものである。

【0085】

【表 2】

表 1

要素	要素タイプ	時間属性
タイトル	AV	式 3
サブセクションN	AV	式 3
図表及び図表注釈N	AV	式 4
要約	V	上記コード 1
参考文献	V	コード 1
図表－注釈なし	V	式 2
ページNのサムネイル	V	TPC
著者名	V	コード 1
公開名	A	式 1
公開日	A	式 1
キーワード	A	式 1
ページ番号	A	式 1
ページ数	A	式 1

これらの図表－注釈なし*は、図表注釈と組み合わせるために選択された図表である。注釈がないか非常に長い図表は、この前処理ではこのグループから除外される。

【0086】

一実施形態では、サムネイルページ定数（TPC）は1秒である。

【0087】

情報コンテンツ属性

文書要素について、情報コンテンツすなわち重要性属性がオーディオ、ビジュアル、及びオーディビジュアル要素に割り当てられる。情報コンテンツ属性を異なる文書要素について計算する。

【0088】

一部の文書要素、例えばタイトルに固定属性を割り当て、一方、その他の文書要素、例えば図表にコンテンツに応じた重要性属性を割り当てることができる。

【0089】

情報コンテンツ属性は、オーディオとビジュアル要素について一定であるか、そのコンテンツから計算される。文書理解やブラウズタスクの場合など、異なるタスクには異なる情報コンテンツ値のセットとしてもよい。これらはアプリケーション制約と考える。

【0090】

文書要素とそのタイプの例を表2に示す。

【0091】

【表 3】

表 2

要素	可能な要素タイプ	好ましい要素タイプ
タイトル	VまたはAV	AV
サブセクションN	VまたはAV	AV
図表及び図表注釈 N	AV	AV
要約	VまたはAV	V
参考文献	VまたはAV	V
図表－注釈なし*	V	V
ページNのサムネ イル	V	V
著者名	VまたはAVまたはA	V
公開名	VまたはAVまたはA	A
公開日	VまたはAVまたはA	A
キーワード	A	A
ページ番号	AまたはAV	A
ページ数	A	A

図表－注釈なし*は、図表注釈と組み合わせるために選択された図表である。注釈がないか注釈が非常に長い図表は、この前処理ではこのグループから除外される。

【0092】

一実施形態では、文書要素をどう表すかユーザのフィードバックに基づき選択する。選択のしかたの例を以下に挙げる：

一実施形態では、テキストを読み出すのにどのくらいかかるかに基づいてVとAVタイプを選択する。短いテキストの場合、オーディオチャンネルがユーザの読む速さにシンクロするので、AVの方がよいと仮定する。要約などの長いテキストの場合、この仮定はあてはまらず、オーディオチャンネルは有用ではなく邪魔な場合もある。

【0093】

公開名と公開日等の非常に短いテキストの場合、オーディオチャンネルをコミュニケーションのために使う。

【0094】

著者名をビジュアル要素とするのは、シンセサイザーが名前の発音を間違えることがあるからである。

【0095】

コンテンツに基づく情報属性の計算

一実施形態では、情報コンテンツ属性を以下のように計算する。

サムネイル

ページ番号に基づく情報コンテンツ（例えば重要性）（最初のページが最も重要であり、それ以降のページでは重要性は減少する）

コンテンツに基づく情報コンテンツ（例えば重要性）（テキストのみ（非常に多くある場合であり、情報コンテンツは大きくない）と、図表入りテキスト（より多くの情報がサムネイルにある））

$$v_{v,i} = \min\{1, 1/\text{pagenum} + PC\}$$

PC：ページの複雑性 PCを [0 0.5] の範囲に規格化する

pagenum: ページ番号

【0096】

【数9】

10

$$PC = \alpha f(\text{title}) + \beta f(\text{sub section}) + \frac{\sum_n a_n w_{e_{v,n}} h_{e_{v,n}}}{w_p h_p} \left\{ \begin{array}{ll} a_n = 1 & \text{page}(e_{v,n}) = \text{pagenum} \\ a_n = 0 & \text{ow} \end{array} \right\} \quad (\text{式10})$$

wはページの幅であり、hはページの高さである。

【0097】

【数10】

20

$$w_{e_{v,n}}, h_{e_{v,n}}$$

はそれぞれビジュアル要素 $e_{v,n}$ の幅と高さである。関数fが文書の情報コンテンツの尺度となる（例えば、エントロピーや特徴などのビジュアルな重要性の尺度であり、これは周知である）。一実施形態では、 α は0.7であり、 β は0.3である。

30

【0098】

図表

同一ページ上の図表のサイズの合計に対する図表サイズ

【0099】

【数11】

$$v_{v,i} = \frac{w_{e_{v,i}} h_{e_{v,i}}}{\sum_n a_n w_{e_{v,n}} h_{e_{v,n}}} \left\{ \begin{array}{ll} a_n = 1 & \text{page}(e_{v,n}) = \text{page}(e_{v,i}) \\ a_n = 0 & \text{ow} \end{array} \right\}. \quad (\text{式11})$$

40

図表の情報コンテンツ属性の計算に役に立つその他の特徴には、図表の複雑性、図表参照（その図表が何回参照されるか）、及び図表コンテンツ（顔や建物が含まれているか）などがある。

【0100】

50

一実施形態における情報属性のサマリー

一実施形態で使用する情報コンテンツ属性のサマリーを表3に示す。

【0101】

【表4】

表 3

要素		情報属性 (v_i)
タイトル	固定	1
サブセクションN	固定	1/40
図表及び図表注釈 N	図サイズ/X	$1/4 * (\text{式 } 1 \ 1)$
要約	固定	1/8-1/24 最初の3つの文のみを示す 2つのタスクに対し、異なる2つの重要性値
参考文献	固定	1/20
図表—注釈なし*	図サイズ/X	$1/4 * (\text{式 } 1 \ 1)$
ページNのサムネ イル	1 / ページ数 + ペー ジ複雑性	式 1 0
著者名	固定	1/90
公開名	固定	1/180
公開日	固定	1/90
キーワード	固定	1/180
ページ番号	固定	1/180
ページ数	固定	1/180

上記の値を使用して、後で修正することもできる。例えば、4頁文書の15－20秒のMMネイルの場合、特定の出力が期待される。期待される出力がこれらの値を用いた出力とマッチしない場合、情報値を修正してもよい。

【0102】

最適化器の例

一実施形態では、ビジュアル情報セグメントとオーディブル情報セグメント、及び出力機器のディスプレイサイズや時間スパンT（最終的なマルチメディアサムネイルの時間的長さ）等の入力に応じて、最適化器が最適化アルゴリズムを実行する。

【0103】

最適化アルゴリズムの主要機能は、各ページを所定時間（例えば0.5秒）だけディスプレイに表示するとして、使える時間スパン中に、ユーザに何ページ表示できるか最初に決定する。

【0104】

一実施形態では、最適化器は、周知のやり方で、リニアパッキング／フィリングオーダーアプローチ (linear packing/filling order approach) をソートした時間属性に対して適用し、マルチメディアサムネイルにどの図表を含めるか選択する。静止画保持を文書の選択された図表に適用する。画像保持によりビジュアルチャネルが使用されている間、注釈をオーディオチャネルで「話す」。最適化後、最適化器は、選択されたビジュアル、オーディオ、及びオーディオビジュアルセグメントを読み出し順序で並べ直す。

【0105】

他の最適化器を用いて、時間スパン L とサイズが制約されたビジュアルディスプレイにおいてともにコミュニケーションされる情報を最大化する。

【0106】

具体的な最適化器の実施

一実施形態では、最適化器は、文書要素の情報コンテンツ (重要性) 属性と時間属性及び時間的制約に基づき、サマリーに含める最良の文書要素を選択するように設計される。

【0107】

より具体的には、問題を以下のように提起することができる：時間制約 T 、各ビジュアル要素 $e_{v,i}$ の重要性 $v_{v,i}$ と時間属性 $t_{v,i}$ 、各オーディオ要素 $e_{a,i}$ の情報コンテンツ $v_{a,i}$ と時間属性 $t_{a,i}$ を与えられると、最適化器は、サムネイルに含めるべきビジュアル要素とオーディオ要素を探す。一部の文書要素 $e_{v,i}$ の場合、 $e_{a,i}$ を含めなければならない (これは、図表に伴う図表注釈等のように、ビジュアル要素と同期しなければならない一部のオーディオ要素の場合である)。他の要素セット (オーディオビジュアル要素 $e_{av,i}$) を加えることにより、この問題を解決し、この要素セットを純粋にビジュアルな要素 (例えばページのサムネイル) 及び純粋なオーディオ要素 (例えばキーワード) から分けることができる。

【0108】

最適化器の一実施形態では、プレゼンテーションに使える時間を最大限使用することにより、ユーザに提示する情報コンテンツを最大化するよう試みる。ビジュアル要素とオーディオビジュアル要素を含める単純化をして、ビジュアル要素が締めている時間にオーディオ要素を重ねれば、最適化問題は「0-1 ナップザック」問題となる。この問題では、要素は完全に含まれるか破棄されるかのいずれかである。これは、R.L.Rivestほか著、「Introduction to Algorithms」、MIT Press、MC-Graw-Hill、Cambridge Massachusetts、1997年に記載されている。

【0109】

以下は具体的な最適化器の実施例である。前処理器が2ページ文書から、以下の情報と時間属性を有する要素を抽出したとする：

タイトル $t=2$, $v=1$ (t : 時間属性、 v : 情報属性)

図1 $t=4$, $v=1/8$ (注釈なし)

図2 $t=10$, $v=1/8$ (注釈あり)

要約 $t=20$, $v=1/8$

キーワード1 $t=1$ $v=1/180$

キーワード2 $t=3$ $v=1/180$

キーワード3 $t=2$ $v=1/180$

一実施形態では、時間制約は $T=8$ である。

【0110】

最初の段階で、ビジュアル要素 (図1、要約) とオーディオビジュアル要素 (タイトル、図2) をそれらの v/t 比を用いて並べる。

【0111】

タイトル ($1/2$)、図1 ($1/32$)、図2 ($1/80$)、要約 ($1/160$)

トータル時間は8である。最初に順序付けリストから、最適化器は、タイトルと図1を選択して表示に含める。トータル時間は $2 + 4 = 6$ である。残り時間 (すなわち $8 - 6 = 2$) にフィットするオーディオビジュアル要素とビジュアル要素はないから、最適化器は

10

20

30

40

50

第２段階に進む。

【０１１２】

第２段階では、オーディブル要素（キーワード）をそれらの v/t 比を用いて以下のよう
に順序づける：キーワード１（１／１８０）、キーワード３（１／３６０）、キーワード２（１／５４０）

ここで、純粋にビジュアルな要素により締められている時間を計算する：

Time（図１）＝４。

【０１１３】

この時間制約を用いてオーディオ要素を挿入する。順序付けられたリストからキーワー
ド１とキーワード３を選択する。

【０１１４】

それゆえ、最終的なマルチメディアサムネイルは次の文書要素を含む：タイトル、図１
、キーワード１及びキーワード３

０－１ナップザック問題－２段階

一実施形態では、最適化器の目標は、以下の条件の下で次の値を最大化することである

【０１１５】

【数１２】

$$\sum_i x_{v,i} v_{v,i} + \sum_i x_{av,i} v_{av,i}$$

20

$$\text{条件：} \quad \sum_i x_{v,i} t_{v,i} + \sum_i x_{av,i} t_{av,i} \leq T$$

$$x_{v,i} \in \{0,1\}$$

$$x_{av,i} \in \{0,1\}$$

30

ここで、

$v_{v,i}$ ：ビジュアル要素 i の情報コンテンツ属性；

$v_{av,i}$ ：オーディオビジュアル要素 i の情報コンテンツ属性；

$t_{v,i}$ ：ビジュアル要素 i の時間制約；

$t_{av,i}$ ：オーディオビジュアル要素 i の時間属性；

T ：時間制約、マルチメディアサムネイル提示の時間的長さ

$x_{v,i}$ 、 $x_{av,i}$ ：変数が含まれているかどうかを示す含み変数。

【０１１６】

次に、最適化器は、比

40

【０１１７】

【数 1 3】

$$\frac{v_{v,i}}{t_{v,i}} \text{ と } \frac{v_{av,i}}{t_{av,i}}$$

10

を、大きい順にソートして、順序付けリスト

【 0 1 1 8】

【数 1 4】

$$L = \{e_{\ell(0)}, \dots, e_{\ell(N-1)}\}$$

20

を作る。一実施形態では、次のステップを反復して順序付けリストを生成する：

【 0 1 1 9】

【数 1 5】

$$T_0 = T$$

$$O = \{\}$$
 空リスト

$$i = 0$$

$$\text{while } (T_0 > 0 \text{ and } i < N) \{$$

$$\quad \text{while } (t_{\ell(i)} > T_0 \text{ and } i < N) \ i++$$

$$\quad \text{if } (i < N)$$

$$\quad \text{add } e_{\ell(i)} \text{ to the output list } O$$

$$T_0 = T_0 - t_{\ell(i)}$$

$$\quad i++$$

$$\quad \}$$

$$\}$$

30

40

ビジュアル文書要素とオーディオビジュアル文書要素のリストを特定した後、最適化器はビジュアル要素を提示するのに要する時間を次のように判断する。

50

【 0 1 2 0 】

【数 1 6】

$$T_v = \sum_i x_{v,i} t_{v,i}, \begin{cases} x_{v,i} = 0 & e_{v,i} \notin 0 \\ x_{v,i} = 1 & e_{v,i} \in 0 \end{cases}$$

10

そして、オーディブル要素の最適化の第 2 段階を実行する：

【 0 1 2 1 】

【数 1 7】

$$\text{最大化} \quad \sum_i x_{a,i} v_{a,i}$$

$$\text{条件：} \quad \sum_i x_{a,i} t_{a,i} \leq T_v$$

20

$$x_{a,i} \in \{0,1\}$$

ここで、

$v_{a,i}$ ：オーディブル要素 i の情報値属性；

$t_{a,i}$ ：オーディブル要素 i の時間属性；

30

$x_{a,i}$ ：含み変数

他の実施形態では、ビジュアル要素の前にオーディブル要素とオーディオビジュアル要素の組み合わせを処理することもできることに留意されたい。最適化器は、ビジュアル要素とオーディブル要素のいずれをオーディオビジュアル要素とともに最初に処理するかを指示する入力に応じる。図 6 に実施例を示す。図 6 を参照して、アプリケーション制約は、2つの段階のいずれを最初に実行するか決定するために段階決定ブロック 6 0 1 により使用される入力である。かかるアプリケーション制約は、例えばオーディオとビジュアルでどちらがより重要かを示す。第 1 段階は、所与の時間制約の下でビジュアル及びオーディオビジュアル情報とオーディオビジュアル情報を最大化し、第 2 段階は、マルチメディア表示に割り当てられるプレゼンテーション時間の残り時間の下でオーディブル情報を最大化する。このように、段階決定ブロックは、最初にどの段階に進み、次にどの段階に進むかを示す。

40

【 0 1 2 2 】

図 7 は、文書の文書要素を選択するプロセスの一実施形態を示すフロー図である。このプロセスは、ハードウェア（例えば回路、専用ロジック等）、（汎用コンピュータシステムまたは専用機上で実行される）ソフトウェア、またはこれらの組み合わせを含む処理ロジックにより実行される。処理ロジックは最適化器の中にあってもよい。

【 0 1 2 3 】

図 7 を参照して、このプロセスの開始において、処理ロジックがオーディブル、ビジュアル、及びオーディオビジュアル要素を処理する（処理ブロック 7 0 1 ）。

50

【0124】

オーディブル、ビジュアル、及びオーディオビジュアル文書要素から、処理ロジックは、ビジュアルとオーディオビジュアル文書要素の組み合わせか、またはオーディブルとオーディオビジュアル文書要素の組み合わせのいずれかを、文書のマルチメディア表示に含めるために、時間属性と情報コンテンツ属性に基づいて選択する（処理ブロック702）。一実施形態では、処理ロジックは、1つ以上のアプリケーション制約に基づいた入力に基づいて、ビジュアルとオーディオビジュアル文書要素の組み合わせか、オーディブル情報のいずれかを選択する。

【0125】

選択後、処理ロジックは、所与の時間制約の下で選択された文書要素中の情報を最大化する（処理ブロック703）。一実施形態では、処理ロジックは、ビジュアルとオーディオビジュアル文書要素を選択し、そのビジュアルとオーディオビジュアル文書要素の情報コンテンツの合計を最大化する一方、選択されたビジュアルとオーディオビジュアル要素に関連する時間属性の合計はマルチメディアプレゼンテーションの長さより短い。すなわち、処理ロジックは要素のクラスを選択する。任意的に、処理ロジックは、文書要素に関連する情報コンテンツ属性の時間属性に対する比に基づいて、ビジュアルとオーディオビジュアル文書要素、またはオーディブルとオーディオビジュアル文書要素を順序付ける。

10

【0126】

その後、選択された文書要素の第1のセットを最大化した後、処理ロジックは、マルチメディア表示を完了するまでに残っている時間量の下で、文書要素の第2のセットを最大化する。一実施形態では、（オーディブル文書要素の方が選択されていれば）ビジュアルとオーディオビジュアル文書要素を選択した後、処理ロジックは、オーディブル文書要素の情報コンテンツの合計を最大化するようにマルチメディア表示に含めるオーディブル要素を選択する。ここで、選択されたオーディブル要素に関連する時間属性の合計は、マルチメディアプレゼンテーションの長さと選択されたビジュアル要素を提示する時間の間の差よりも小さい。

20

【0127】

ページサムネイル制約を伴う「ナップザック問題」

一実施形態では、ユーザによりよい状況を与え、文書中をより容易にナビゲーションできるように、ページに表示すべきビジュアルアイテムがある場合、そのページのサムネイルビューを含める。それゆえ、そのページが以前に加えられていない場合、そのページを加えるコストがかかるであろう。そのコストはページサムネイルを加えることによるものである。よって、新しい制約を加え、次のように表す：

30

【0128】

【数 1 8】

$$\text{最大化} \quad \sum_i x_{v,i} v_{v,i} + \sum_i x_{av,i} v_{av,i}$$

$$\text{条件:} \quad \sum_i x_{v,i} t_{v,i} + \sum_i x_{av,i} t_{av,i} \leq T$$

$$x_{v,i} \in \{0,1\}$$

$$x_{av,i} \in \{0,1\}$$

ここで、

$$\begin{aligned} \text{及び} \quad & x_{v,i} = 1 \Rightarrow x_{v, \arg \text{page}(e_{v,i})} = 1 \\ & x_{av,i} = 1 \Rightarrow x_{v, \arg \text{page}(e_{av,i})} = 1 \end{aligned}$$

10

20

ここで、 $\text{page}(e_{v,i})$ は、要素 $e_{v,i}$ があるページのサムネイルを含む要素 $e_{v,j}$ であり、 $\text{page}(e_{av,i})$ は、要素 $e_{av,i}$ があるページのサムネイルを含む要素 $e_{v,j}$ である。

【0 1 2 9】

このプロセスは、上記のプロセス、すなわち処理ロジックが比

【0 1 3 0】

【数 1 9】

$$\frac{v_{v,i}}{t_{v,i}} \text{ と } \frac{v_{av,i}}{t_{av,i}}$$

30

とを大きい順にソートして、要素数 N の順序リスト $L=\{e_{l(0)}, \dots, e_{l(N-1)}\}$ を形成するものと同様である。最適化器は次のステップを反復する：

【0 1 3 1】

【表 5】

```

 $T_0 = T$ 
 $O = \{\}$  空リスト
 $P = \{\}$  空リスト
 $TN = \{e_{v,n} \dots\}$  サムネイルであるビジュアル要素のリスト
 $t_{TN} = 0.5$ 
 $i = 0$ 
while ( $T_0 > 0$  and  $i < N$ ) {
    addpage =  $e_{\ell(i)} \notin TN$  and  $page(e_{\ell(i)}) \notin P$ 
    while (!((addpage and ( $t_{\ell(i)} + t_{TN} \leq T_0$ )) or (!addpage and ( $t_{\ell(i)} \leq T_0$ ))) and ( $i < N$ ))
    {
        i++
        if ( $i < N$ )
            addpage =  $e_{\ell(i)} \notin TN$  and  $page(e_{\ell(i)}) \notin P$ 
    }
    if ( $i < N$ )
        if (addpage)
        {
            add page =  $e_{\ell(i)}$  to  $P$ 
             $T_0 = T_0 - t_{\ell(i)} - t_{TN}$ 
            add  $e_{v,j} = page(e_{\ell(i)})$  to the output list  $O$ 
            remove  $e_{v,j} = page(e_{\ell(i)})$  from  $L$ 
             $N--$ 
        }
    else
    {
         $T_0 = T_0 - t_{\ell(i)}$ 
    }
    add  $e_{\ell(i)}$  to the output list  $O$ 
    i++
}

```

オーディオ、ビジュアル、及びオーディオビジュアル要素の結合最適化器－１段階

アプリケーションにより、ビジュアルまたはオーディオ要素の配置に関するプライオリ 50

ティが与えられない方が望ましいのであれば、最適化問題は次のように定式化できる：

【0 1 3 2】

【数 2 0】

$$\text{最大化} \quad \sum_i x_{v,i} v_{v,i} + \sum_i x_{av,i} v_{av,i} + \sum_i x_{a,i} v_{a,i}$$

$$\text{条件：} \quad \sum_i x_{av,i} t_{av,i} + \max \left\{ \sum_i x_{v,i} t_{v,i}, \sum_i x_{a,i} t_{a,i} \right\} \leq T$$

10

$$0 \leq x_{v,i} \leq 1$$

$$0 \leq x_{av,i} \leq 1$$

$$0 \leq x_{a,i} \leq 1$$

この問題は次の問題と等価である：

20

【0 1 3 3】

【数 2 1】

$$\text{最大化} \quad \sum_i x_{v,i} v_{v,i} + \sum_i x_{av,i} v_{av,i} + \sum_i x_{a,i} v_{a,i}$$

$$\text{条件：} \quad \begin{aligned} \sum_i x_{av,i} t_{av,i} + \sum_i x_{v,i} t_{v,i} &\leq T \\ \sum_i x_{av,i} t_{av,i} + \sum_i x_{a,i} t_{a,i} &\leq T \end{aligned}$$

30

$$0 \leq x_{v,i} \leq 1$$

$$0 \leq x_{av,i} \leq 1$$

$$0 \leq x_{a,i} \leq 1$$

一般的に、これは線形問題であり、標準的なLP問題解決プログラムを使用することができる。例えば、Matlab(<http://www.mathworks.com>)やCVSoptソフトウェアパッケージ(<http://www.ee.ucla.edu/~vandenbe/cvxopt/>)に含まれているものである。

40

【0 1 3 4】

要素 e_i の時間属性 t_i は2つの部分 t_{ia} と t_{ib} に($t_{ia}+t_{ib}=t_i$)分けることができる。ここで、 t_{ia} は時間 t_i の大部分であり、 t_{ib} は残り少しである($t_{ia}>t_{ib}$)。最適化問題を次のように定式化する：

最大化

【0 1 3 5】

【数 2 2】

$$\sum_i x_{v,ia} v_{v,i} + \sum_i x_{av,ia} v_{av,i} + \sum_i x_{a,ia} v_{a,i} + \sum_i x_{v,ib} f(v_{v,i}) + \sum_i x_{av,ib} f(v_{av,i}) + \sum_i x_{a,ib} f(v_{a,i})$$

条件：

$$\sum_i x_{av,ia} t_{av,ia} + \max \left\{ \sum_i x_{v,ia} t_{av,ia}, \sum_i x_{a,ia} t_{a,ia} \right\} + \sum_i x_{av,ib} t_{av,ib} + \max \left\{ \sum_i x_{v,ib} t_{av,ib}, \sum_i x_{a,ib} t_{a,ib} \right\} \leq T$$

10

$$0 \leq x_{v,ia} \leq 1$$

$$0 \leq x_{av,ia} \leq 1$$

$$0 \leq x_{a,ia} \leq 1$$

$$0 \leq x_{v,ib} \leq 1$$

$$0 \leq x_{av,ib} \leq 1$$

$$0 \leq x_{a,ib} \leq 1$$

ここで、f は

20

$$f(v_{q,i}) = \alpha_{q,i} v_{q,i} \quad \text{with} \quad \alpha_{q,i} = \frac{t_{q,ib}}{t_{q,ia}} \quad \text{or} \quad \alpha_{q,i} = \frac{t_{q,ib}}{t_{q,i}}$$

q=v, av, または a の場合

値 $\alpha_{q,i}$ はユーザまたはアプリケーションが $0 < \alpha_{q,i} < t_{ib}/t_{ia}$ を満たす定数として選択することができる。

30

【0 1 3 6】

時間属性を主要部分と残差部分に分け、主要部分（であるが要素の時間の全部ではないもの）を最終的な選択に含めることにより、文書要素の選択プロセスをよりフレキシブルにすることができる。

【0 1 3 7】

合成

上記の通り、最適化器は分析機から出力を受け取る。この出力は、ビジュアルとオーディブル文書情報の特徴と、装置の特徴と、または1つ以上の制約（例えば、ディスプレイサイズ、使える時間スパン、ユーザ設定の好み、装置のパワー容量）を含む。そして、最適化器は、装置の制約に合ったビジュアルとオーディブル情報の組み合わせを計算し、利用できる出力ビジュアル及びオーディオチャネルを介して出力できる情報キャパシティを利用する。このように、最適化器は選択機または選択メカニズムとして動作する。

40

【0 1 3 8】

選択後、シンセサイザー（synthesizer）（合成機）が最終的なマルチメディアサムネイルを作成する。一実施形態では、シンセサイザーが最適化器で決定された選択されたマルチメディア処理ステップを実行して、最終的マルチメディアサムネイルを作成する。一実施形態では、シンセサイザーは、処理ステップをリストしたプレーンテキストファイルやXMLファイル等のファイルを受け取る。他の実施形態では、処理ステップのリストは、例えば、2つのソフトウェアモジュール間のソケット通信またはCOMオブジェクト通

50

信等の手段を介して、シンセサイザーに送られる。さらに他の実施形態では、処理ステップのリストは、両方のモジュールが同一ソフトウェア中にある場合、関数パラメータとして渡される。マルチメディア処理ステップは、「従来の」画像処理ステップであるクロッピング、スケーリング、ペーストを含んでもよいが、時間成分を含むステップである例えばページフリッピング、パン、ズーム、及び音声・音楽合成等を含んでもよい。

【0139】

一実施形態では、シンセサイザーは、ビジュアルシンセサイザー、オーディオシンセサイザー、及びシンクロナイザ／コンポーザを含む。シンセサイザーは、ビジュアル合成を用いて選択されたビジュアル情報から画像または画像のシーケンスを合成し、オーディオシンセサイザーを用いてオーディブル情報から音声を合成し、次にシンクロナイザ／コン

10

【0140】

一実施形態では、ズームとページフリッピング等の（オーディオを伴わない）画像シーケンスを含むビジュアル文書は、Adobe社のAfterEffectsを用いて実行され、一方、シンクロナイザ／コンポーザとしてAdobePremierを使用する。一実施形態では、オーディオシンセサイザーとしては、CMU音声合成ソフトウェア（FestVox,<http://festvox.org/voicedemos.html>）を用いて、オーディブル情報の音声を生成する。

20

【0141】

一実施形態では、シンセサイザーは、シンクロナイザ／コンポーザを含まない。このような場合、シンセサイザーの出力は、2つの別々のストリームとして出力される。1つはオーディオであり、1つはビジュアルである。

【0142】

シンクロナイザ／コンポーザの出力は、単一ファイルに結合されてもよいし、オーディオとビデオチャンネルに分けてもよい。

【0143】

アプリケーション

上記の方法は、多数のアプリケーションに潜在的に有用である。例えば、この方法は、モバイル装置や多機能機（MFP）等の装置で文書をブラウズするために使用してもよい。

30

【0144】

例えば、モバイル装置上でインターラクティブに文書ブラウズを実行する時、文書ブラウズは、ズームやスクロール等ではなく、プレイ、ポーズ、早送り、加速、減速等を含む動作として再定義することができる。

【0145】

他のモバイル装置アプリケーションでは、モバイル装置で文書を見るとき、上記の方法を使用して、MMネイルの長いもの（例えば15分の長さ）を使用し、概要だけではなく、文書の内容を理解させるようにすることもできる。このアプリケーションは、画像表示機能が限定されているが、オーディオ機能が充実している装置（例えば携帯電話）にも適している。一実施形態では、モバイル装置で文書をブラウズした後、そのモバイル装置はそれを他の場所にある装置（例えばMFP）に送り、その装置にその文書に対して他の処理を実行させる（例えば、その文書を印刷させる）。

40

【0146】

MFPの一アプリケーションでは、ここで説明した方法を用いて文書の概要を知ることができる。例えば、ユーザがMFPで文書をコピーしている時、ページをスキャンすると、自動的に計算された文書の概要がユーザに表示され、その文書の内容を理解しやすくなる。

【0147】

50

MFP中で文書画像の画質向上を実行する画像処理アルゴリズムは、問題となりそうな領域、例えばコントラストが低い領域、フォントが小さい領域、スキャンの分解能と干渉してしまうハーフトーン領域等を検出する。ユーザにスキャン文書の品質（すなわちスキャン品質）を評価させるために、MMネイルをコピー機のディスプレイ上に（オーディオなしで）表示して、必要に応じて設定変更（例えば、コントラストを高くしたり、解像度を上げたりする）を示唆する。

【0148】

翻訳アプリケーションでは、オーディオチャネルの言語をユーザが選択して、オーディブル情報をその選択された言語で出力してもよい。この場合、最適化器は言語が異なれば異なる機能をするが、それはオーディオの長さが異なるからである。すなわち、最適化器の結果は言語によって異なる。一実施形態では、ビジュアル文書テキストを変更する。ビジュアル文書部分を異なる言語で再表示することができる。

10

【0149】

一実施形態では、MMネイルの最適化は、ユーザからのインタラクションに基づいて、すぐに計算される。例えば、ユーザがオーディオチャネルを閉じると、この情報チャネルがなくなったことに対応して、他のビジュアル情報によりビジュアルの表示が変化する。他の実施例では、ユーザがビジュアルチャネルをスローダウンすると（例えば、車の運転中）、オーディオチャネルを介して出力される情報が変化する。また、ビューイング装置の計算上の制約に基づいて、例えばズーム、パン等のアニメーション効果がえられる。

【0150】

20

一実施形態では、MMネイルを使用して、障害を持つ人が文書情報を得るのを援助する。例えば、目が不自由な人は、短いテキストをオーディブル情報の形で得たいかも知れない。他の実施例では、色盲の人は、文書中の色のついた情報の一部をオーディオチャネルのオーディブル情報として得ることを欲するかも知れない。

【0151】

コンピュータシステムの実施例

図8は、ここに記載した1つ以上の動作を実行するコンピュータシステムの例を示すブロック図である。図8を参照して、コンピュータシステム800は、クライアントまたはサーバのコンピュータシステムを含む。コンピュータシステム800は、情報をやりとりする通信メカニズムすなわちバス811と、情報を処理する、バス811に結合したプロセッサ812とを有する。プロセッサ812は、例えばペンティアム（登録商標）プロセッサ等のマイクロプロセッサを含むが、マイクロプロセッサに限定されない。

30

【0152】

システム800は、さらに、プロセッサ812により実行される情報及び命令を格納する、バス811に結合したランダムアクセスメモリ（RAM）またはその他のダイナミック記憶装置804（ここではメインメモリと呼ぶ）を有する。メインメモリ804は、プロセッサ812による命令の実行中に、一時的変数やその他の中間情報を記憶するために使用される。

【0153】

コンピュータシステム800は、プロセッサ812の静的情報や命令を記憶する、バス811に結合した読み出し専用メモリ（ROM）及び／またはその他の静的記憶装置806と、磁気ディスク、光ディスクとその対応するディスクドライブ等であるデータ記憶装置807とを有する。データ記憶装置807は、情報と命令を記憶し、バス811に結合している。

40

【0154】

コンピュータシステム800は、コンピュータのユーザに情報を表示するための、バス811に結合した、陰極線管（CRT）または液晶ディスプレイ（LCD）等のディスプレイ装置821に結合している。英数字入力装置822は、英数字その他のキーを含み、バス811に結合され、プロセッサ812に情報とコマンド選択を送る。追加的なユーザ入力装置として、マウス、トラックボール、トラックパッド、スタイラス、またはカーソ

50

ル方向キー等のカーソル制御 8 2 3 があり、バス 8 1 1 に結合し、プロセッサ 8 1 2 に方向情報とコマンド選択を送り、ディスプレイ 8 2 1 条のカーソルの動きを制御する。

【0 1 5 5】

バス 8 1 1 に結合した他の装置としてハードコピー装置 8 2 4 がある。このハードコピー装置 8 2 4 は、紙、フィルム、その他のメディア上に、命令、データ、その他の情報を印刷するために使用される。さらに、スピーカ及び／またはマイクロホン等の音声録音再生装置も任意的にバス 8 1 1 に結合しており、コンピュータシステム 8 0 0 のオーディオインターフェイスとして機能する。バス 8 1 1 に結合する他の装置として、電話やハンドヘルドパームトップ装置と通信する、有線または無線の通信機能 8 2 5 がある。

【0 1 5 6】

なお、オーディブル、ビジュアル、及びオーディオビジュアル文書要素は、図 8 のコンピュータシステムにおける 8 0 6 や 8 0 7 に記録されている各文書要素を読み出すことで各文書要素を受け取ることとしても良いし、8 2 5 によりネットワークを介して接続される他のコンピュータシステムから各文書要素を受信すること、受け取ることとしても良い。

【0 1 5 7】

ディスプレイの制約は、図 8 のコンピュータシステムにおける 8 2 1 から、事前に、又は、処理ロジックの実行時に、ディスプレイ制約を受信して 8 0 6 や 8 0 7 に記録しておくこととしても良い。8 2 5 によりネットワークを介して接続される他のコンピュータシステムにおけるディスプレイのディスプレイ制約の受信についても同様に、事前に、又は、

【0 1 5 8】

システム 8 0 0 のどの構成要素もそれに関連するハードウェアも、本発明で使用してもよい。しかし、言うまでもなく、他の構成のコンピュータシステムでは、これらの構成要素の一部または全部を含んでもよい。

【0 1 5 9】

上記の説明を読んだ当業者には本発明の変形例や修正例が明らかになったことは間違いなく、言うまでもなく、上記のどの実施形態も本発明を限定することを目的としたものではない。それゆえ、いろいろな実施形態の詳細の説明は、本発明に本質的であると考えられる特徴のみを記載した請求項の範囲を限定するものではない。

【0 1 6 0】

なお、本開示にあたり、以下の付記を記す。

(付記 2 4)

システムにより実行された時、システムに一方法を実行させる命令を格納した 1 つ以上の記録媒体を有する製品であって、前記方法は、オーディブル、ビジュアル、及びオーディオビジュアル文書要素を受け取るステップと、マルチメディア表示に含める 1 つ以上のオーディブル、ビジュアル、オーディオビジュアル文書要素のセットを、ディスプレイと時間制約、及び少なくとも 1 つのアプリケーション制約とユーザごとの制約に基づいて選択することを特徴とする製品。

(付記 2 5)

付記 2 4 に記載の製品であって、

1 つ以上のオーディオ、ビジュアル、及びオーディオビジュアル文書要素のセットを選択するステップは、

ディスプレイ制約、及びアプリケーション制約とユーザ制約の少なくとも一方に基づき、オーディブル、ビジュアル、及びオーディオビジュアル文書要素と関連する時間属性を計算するステップと、

ディスプレイとアプリケーションの制約に基づき、オーディブル、ビジュアル、及びオーディオビジュアル文書要素と関連する情報コンテンツ属性を計算するステップと、を有することを特徴とする製品。

(付記 2 6)

付記 24 に記載の製品であって、

1 つ以上のオーディオ、ビジュアル、及びオーディオビジュアル文書要素のセットを選択するステップは、

前記 1 つ以上のオーディオ、ビジュアル、及びオーディオビジュアル文書要素のセットを用いて前記マルチメディア表示を作成するステップと、を有することを特徴とする製品。

(付記 27)

オーディブル、ビジュアル、及びオーディオビジュアル文書要素を受け取る入力と、マルチメディア表示に含める 1 つ以上のオーディブル、ビジュアル、オーディオビジュアル文書要素のセットを、ディスプレイと時間制約、及び少なくとも 1 つのアプリケーション制約とユーザごとの制約に基づいて選択する選択器とを有することを特徴とする装置。

10

(付記 28)

付記 27 に記載の装置であって、

マルチメディア表示発生器は、1 つ以上のオーディブル、ビジュアル、及びオーディオビジュアル文書要素のセットを用いて前記マルチメディア表示を作成することを特徴とする装置。

(付記 29)

文書のオーディオビジュアル文書要素を受け取るステップと、

オーディオビジュアル文書要素のテキスト部分をビジュアルにナビゲーションするのに必要な時間と、オーディオビジュアル文書要素に関連する合成テキストを再生するのに要する時間のうち最大のものとしてオーディオビジュアル文書要素の時間属性を計算するステップと、を有することを特徴とする方法。

20

(付記 30)

文書のオーディオビジュアル文書要素を受け取るステップと、

オーディオビジュアル文書要素の図表部分をビジュアルに理解するのに必要な時間と、オーディオビジュアル文書要素に関連する合成テキストを再生するのに必要な時間のうち最大のものとしてオーディオビジュアル文書要素の時間属性を計算するステップと、を有することを特徴とする方法。

(付記 31)

文書のビジュアル文書要素を受け取るステップと、

ビジュアル要素の時間属性を計算するステップと、を有し、その計算は、

30

必要なパンとズームの動作を決定するステップと、

その動作が実行される 1 つ以上の解像度に基づいて動作を実行するのに必要な時間を決定するステップと、を有することを特徴とする方法。

(付記 32)

付記 31 に記載の方法であって、

ビジュアルテキスト要素の時間属性の計算は、さらに、

ディスプレイとアプリケーション制約と文書情報に基づき第 1 のビューのビジュアルテキスト要素の可読性を決定するステップと、

ビジュアル要素が可読であると判断された場合、音声合成定数とビジュアル要素中の文字数の積に時間属性を設定するステップと、を有することを特徴とする方法。

40

(付記 33)

付記 32 に記載の方法であって、

第 1 のビューで可読でない場合、

ズームビューでのビジュアル要素の可読性を決定するステップと、

ズームビューで可読であれば、音声合成定数とビジュアル要素中の文字数の積と、ズームビュー中のビジュアル要素中のタイトルにズームするのに必要な時間を示す定数とを足し合わせた値に、時間制約を設定するステップと、を有することを特徴とする方法。

(付記 34)

付記 33 に記載の方法であって、

第 1 またはズームビューで可読でない場合、

50

ビジュアル要素中のテキストの可読性に必要なズーム量を示すズームファクタを計算するステップと、
パンに必要な時間を決定するステップと、
パンピクセル制約と、時間属性を1ラインの元のテキスト幅及び高さでディスプレイ幅の間の差との積と、
ズームビューにおいてビジュアル要素中のタイトルにズームするのに必要な時間を示す定数と、
音声合成定数と、ディスプレイ高さの最小可読高さとの商をビジュアルテキスト要素の文字数から引いた結果との積と、を加えた値に設定することを特徴とする方法。

(付記35)

10

付記32に記載の方法であって、
ディスプレイ制約はディスプレイの幅と高さを含み、文書情報はページの幅と高さに基づくことを特徴とする方法。

(付記36)

付記32に記載の方法であって、
第1のビューはサムネイルビューを含むことを特徴とする方法。

(付記37)

システムにより実行された時、システムに一方法を実行させる命令を格納した1つ以上の記録媒体を有する製品であって、前記方法は、
文書のビジュアルテキスト文書要素を受け取るステップと、
ビジュアルテキスト要素の時間属性を計算するステップと、を有し、その計算は、
必要なパンとズームの動作を決定するステップと、
その動作が実行される1つ以上の解像度に基づいて動作を実行するのに必要な時間を決定するステップと、を有することを特徴とする方法。

20

(付記38)

文書のビジュアル文書要素を受け取るステップと、
ビジュアル要素の時間属性を計算するステップと、を有し、その計算は、
最小可読文字高さへのズーム動作を実行するステップと、
要素の幅をチェックしてその要素がディスプレイ幅にフィットするか決定するステップと、
要素幅がディスプレイ幅にフィットしない場合パン動作を実行するステップと、を有することを特徴とする方法。

30

(付記39)

文書のビジュアル文書要素を受け取るステップと、
ビジュアル要素の時間属性を計算するステップと、を有し、その計算は、
第1のビューにズーム動作を実行するステップと、
要素が完全に見えなければパンを実行するステップと、を有することを特徴とする方法。

(付記40)

オーディブル、ビジュアル、及びオーディオビジュアル文書要素を受け取るステップと、
時間と情報コンテンツ属性に基づき、文書のマルチメディア表示に含めるビジュアルとオーディオビジュアル文書要素の組み合わせかオーディブルとオーディオビジュアル文書要素の組み合わせのいずれかを選択するステップと、
所与の時間制約の下でセットに含まれる情報を最大化するように第1のセットを選択するステップと、
選択された文書要素を最大化した後、マルチメディア表示の残り時間量の下で、セットに含まれる情報を最大化するように第2のセットを選択するステップと、を有することを特徴とする方法。

40

(付記41)

付記40に記載の方法であって、
組み合わせとオーディブル情報のいずれかの選択は、1つ以上のアプリケーション制約に

50

基づいた入力に基づくことを特徴とする方法。

(付記 4 2)

付記 4 0 に記載の方法であって、

選択した文書要素の最大化はページサムネイル制約の下で行われることを特徴とする方法。

(付記 4 3)

付記 4 0 に記載の方法であって、

文書要素に関連する情報コンテンツ属性の文書要素に関連する時間属性に対する比に基づいて、ビジュアルとオーディオビジュアル文書要素を順序付けるステップと、

マルチメディア表示に含めるためのビジュアルとオーディオビジュアル文書要素の組み合わせかオーディブル文書要素のいずれかを選択するステップは、ビジュアルとオーディオビジュアル文書要素の情報コンテンツの合計を最大化し、選択されたビジュアル及びオーディオビジュアル要素に関連付けられた時間属性の合計はマルチメディア表示の表示時間よりも短いように、ビジュアルとオーディオビジュアル文書要素を選択するステップを含むことを特徴とする方法。

10

(付記 4 4)

付記 4 3 に記載の方法であって、

マルチメディア表示に含めるためにビジュアルとオーディオビジュアル文書要素の組み合わせかオーディブル文書要素のいずれかを選択するステップは、オーディブル文書要素の情報コンテンツの合計を最大化し、選択されたオーディブル要素と関連する時間属性の合計がマルチメディア表示の長さで選択されたビジュアル要素の表示時間の間の差よりも小さいように、マルチメディア表示に含ませるオーディオ要素を選択するステップを有することを特徴とする方法。

20

(付記 4 5)

システムにより実行された時、システムに一方法を実行させる命令を格納した 1 つ以上の記録媒体を有する製品であって、前記方法は、

オーディブル、ビジュアル、及びオーディオビジュアル文書要素を受け取るステップと、時間と情報コンテンツ属性に基づき、文書のマルチメディア表示に含めるビジュアルとオーディオビジュアル文書要素の組み合わせかオーディブル文書要素のいずれかを選択するステップと、

30

所与の時間制約の下で文書要素の第 1 のセットに含まれる情報を最大化するように第 1 のセットを選択するステップと、

選択された文書要素を最大化した後、マルチメディア表示の残り時間量の下で、文書要素の第 2 のセットに含まれる情報を最大化するように第 2 のセットを選択するステップと、を有することを特徴とする方法。

(付記 4 6)

オーディオとビジュアル要素を受け取るステップと、

オーディオとビジュアルフォーカスポイントを決定するステップと、

オーディオビジュアル要素を選択するステップと、

選択されたオーディオビジュアル要素と、選択されたオーディオビジュアル要素の一部ではないオーディオとビジュアル要素の組み合わせとに基づき、制約されたディスプレイ領域中のプリセット時間中に表示するナビゲーションパスを生成するステップと、を有することを特徴とする方法。

40

(付記 4 7)

オーディオとビジュアル要素を受け取る入力と、

オーディオとビジュアルフォーカスポイントを決定するオーディオ及びビジュアルフォーカス決定器と、

オーディオビジュアル要素を選択するオーディオビジュアル要素選択器と、

選択されたオーディオビジュアル要素と、選択されたオーディオビジュアル要素の一部ではないオーディオとビジュアル要素の組み合わせとに基づき、制約されたディスプレイ領

50

域中のプリセット時間中に表示するナビゲーションパスを生成するナビゲーションパス生成器と、を有することを特徴とする装置。

【図面の簡単な説明】

【0161】

【図1】文書のマルチメディア表示を生成するプロセスの一実施形態を示すフロー図である。

【図2】ビジュアル情報、可聴情報、オーディオビジュアル情報を処理するプロセスの一実施形態を示すフロー図である。

【図3】文書のビジュアルテキスト文書要素の時間属性を計算するプロセスの一実施形態を示すフロー図である。

10

【図4】文書のビジュアルテキスト文書要素の時間属性を計算するプロセスの他の実施形態を示すフロー図である。

【図5】文書のオーディオビジュアル文書要素の時間属性を生成するプロセスの一実施形態を示すフロー図である。

【図6】2段階最適化器の一実施形態を示す図である。

【図7】文書の文書要素を選択するプロセスの一実施形態を示すフロー図である。

【図8】コンピュータシステムの一実施形態を示すブロック図である。

【符号の説明】

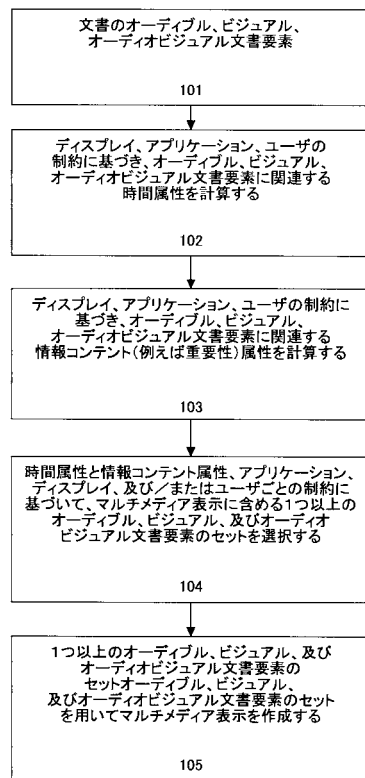
【0162】

- 201 オーディブル文書要素
- 202 ビジュアル文書要素
- 203 オーディオビジュアル文書要素
- 204／205 時間属性と重要性属性を計算
- 206 最適化器
- 207 ナビゲートパス
- 210 ディスプレイ制約
- 211 アプリケーション／タスク制約
- 212 時間制約
- 804 メインメモリ
- 806 スタティック記憶装置
- 807 データ記憶装置
- 812 プロセッサ
- 811 バス
- 821 ディスプレイ装置
- 822 英数字入力装置
- 823 カーソル制御装置
- 824 ハードコピー装置
- 825 通信機能

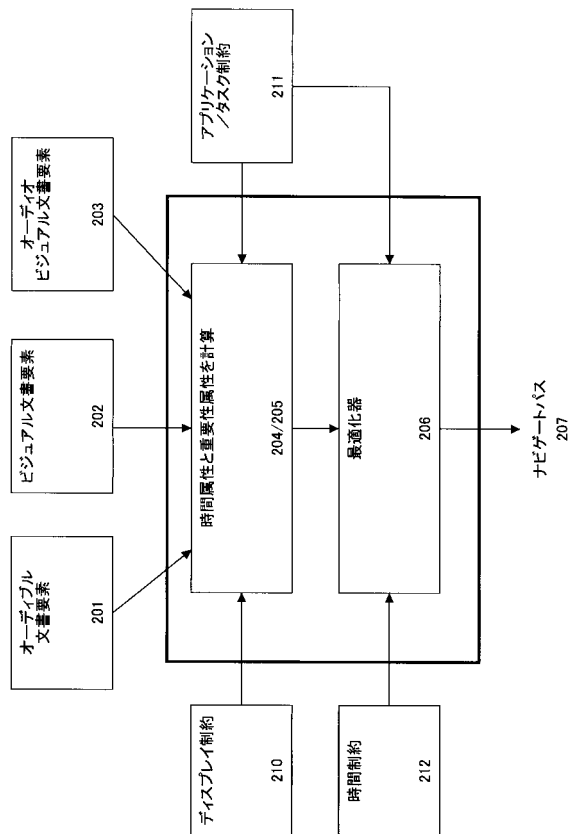
20

30

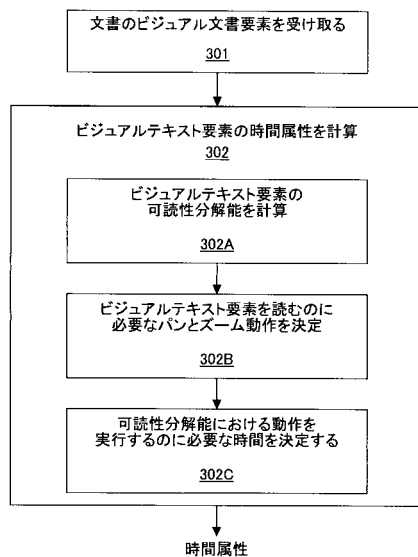
【図 1】



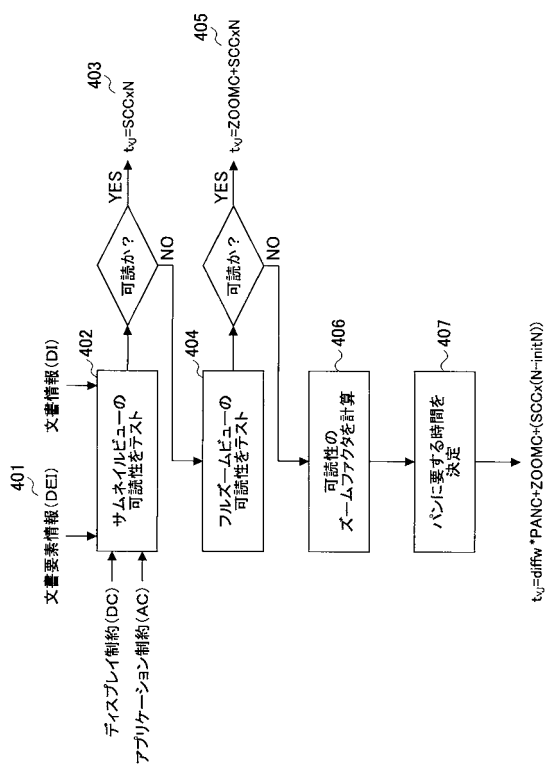
【図 2】



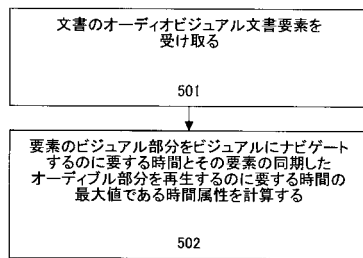
【図 3】



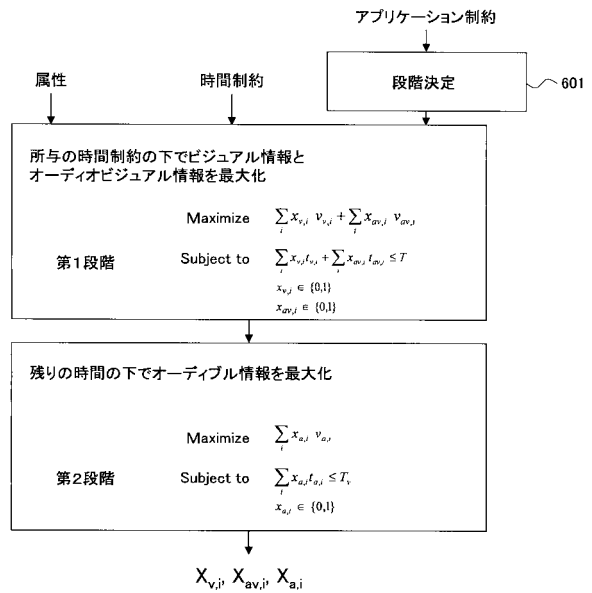
【図 4】



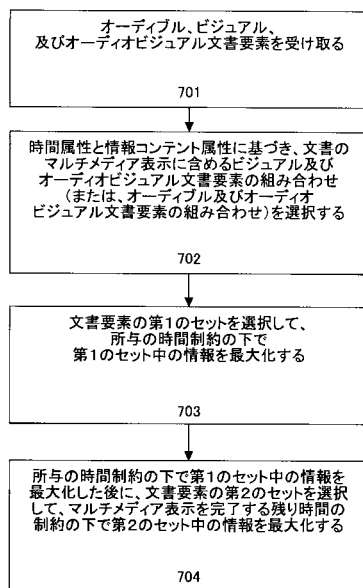
【図 5】



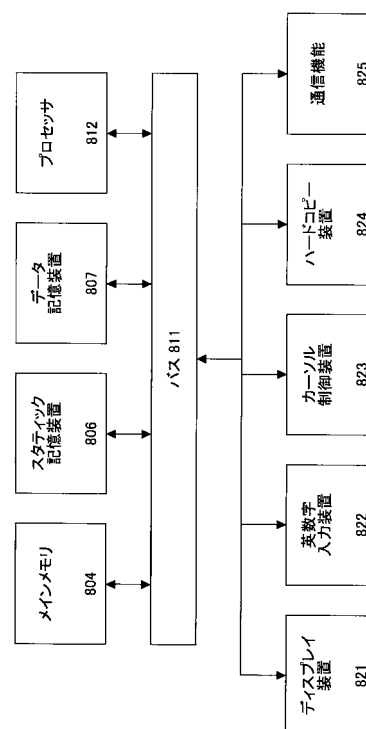
【図 6】



【図 7】



【図 8】



フロントページの続き

- (72)発明者 キャサリン パークナー
アメリカ合衆国, カリフォルニア 94025, メンロ・パーク, サンド・ヒル・ロード 288
2番, スイート 115 リコー イノベーション インク内
- (72)発明者 シドハース ジョシ
アメリカ合衆国, カリフォルニア 94025, メンロ・パーク, サンド・ヒル・ロード 288
2番, スイート 115 リコー イノベーション インク内
- (72)発明者 エドワード シュワルツ
アメリカ合衆国, カリフォルニア 94025, メンロ・パーク, サンド・ヒル・ロード 288
2番, スイート 115 リコー イノベーション インク内

審査官 森田 充功

- (56)参考文献 特開平06-231187 (JP, A)
特開平06-332893 (JP, A)
特開平08-050471 (JP, A)
特開平08-235219 (JP, A)
特開平10-039846 (JP, A)
特開2000-076473 (JP, A)
特開2000-231475 (JP, A)
特開2002-182815 (JP, A)
特開2005-251034 (JP, A)
特開2005-293174 (JP, A)
米国特許出願公開第2005/0084136 (US, A1)

(58)調査した分野(Int.Cl., DB名)

G06F 3/14
G06F 3/048
G06F 3/16