

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号

特許第7199756号

(P7199756)

(45)発行日 令和5年1月6日(2023.1.6)

(24)登録日 令和4年12月23日(2022.12.23)

(51)国際特許分類

H 0 4 N 21/84 (2011.01)

F I

H 0 4 N 21/84

請求項の数 7 (全14頁)

(21)出願番号	特願2021-520997(P2021-520997)	(73)特許権者	505224569
(86)(22)出願日	平成30年12月28日(2018.12.28)		インハ インダストリー パートナーシッ プ インスティテュート
(65)公表番号	特表2022-505092(P2022-505092 A)		Inha-Industry Part nership Institute
(43)公表日	令和4年1月14日(2022.1.14)		大韓民国 2 2 2 1 2 インチョン ミチ ヨルグ インハーロ 1 0 0
(86)国際出願番号	PCT/KR2018/016854	(74)代理人	110001519
(87)国際公開番号	WO2020/080606		弁理士法人太陽国際特許事務所
(87)国際公開日	令和2年4月23日(2020.4.23)	(72)発明者	チョ、クン シク
審査請求日	令和3年4月15日(2021.4.15)		大韓民国 2 1 9 9 6 インチョン ヨン スグ ヘドゥイーロ 8 4ボンギル 1 0、ナンバー 6 0 1-1 3 0 2
(31)優先権主張番号	10-2018-0124863	審査官	富樫 明
(32)優先日	平成30年10月19日(2018.10.19)		
(33)優先権主張国・地域又は機関	韓国(KR)		

最終頁に続く

(54)【発明の名称】 ビデオメタデータと台本データを活用したビデオコンテンツ統合メタデータ自動生成方法
およびシステム

(57)【特許請求の範囲】

【請求項1】

台本処理部と互いに連結するためにビデオの入力を受けてビデオのショットおよびシー
ンを認識して整列する前処理過程を実行し、ビデオコンテンツを分析するビデオ処理部、

ビデオ処理部と互いに連結するために台本の入力を受けて台本およびシーンを認識して
整列する前処理過程を実行し、台本コンテンツを分析する台本処理部、

前記ビデオ処理部で生成されたアノテーションメタデータを格納するアノテーション知
識ベース、および

前記台本処理部で生成されたナラティブメタデータを格納するナラティブ知識ベース
を含み、

アノテーション知識ベースとナラティブ知識ベースは、

エピソード、シーケンス、シーン、およびショットを共有することで相互補完的に使用
してメタデータを自動で生成する、

メタデータ自動生成システムであって、

台本処理部は、

半構造化された台本書式を分析し、台本から人物、発話、説明文、場所、時間を区分す
るために多様な形態の台本構造様式と入力された台本構造とを比較して適した台本様式を
選択し、構造化(Structured)し、該当の様式が入力された台本によるもので
あるかを検証し、誤入力された様式が構造化結果として変形した場合には補正をし、台本
の構造から台本のシーンを検出して情報をメタデータの階層構造として指定する、

メタデータ自動生成システム。

【請求項 2】

ビデオ処理部は、

入力されたビデオの毎フレーム間の類似度を測定し、類似度が予め定められた基準以上に急変するフレーム前後でビデオのショットを検出して区分し、区分されたビデオのすべてのショットで時空間が変わる時点を認識してビデオのシーンを検出する、

請求項 1 に記載のメタデータ自動生成システム。

【請求項 3】

ビデオと台本のシーンの順序を一致させるために、ビデオのあるシーンの発話を音声認識技術によって字幕化して台本の発話と文字列比較アルゴリズムを利用して 1 次整列した後、ダイナミックプログラミング技法を利用してビデオと台本のシーンの順序を整列する、

請求項 1 に記載のメタデータ自動生成システム。

【請求項 4】

台本をビデオのショット単位まで一致させるために、ビデオのシーン内に属するショット内の時間および音声テキスト情報を台本コンテンツ分析過程に伝達して処理する、

請求項 3 に記載のメタデータ自動生成システム。

【請求項 5】

台本処理部と互いに連結するためにビデオの入力を受けてビデオのショットおよびシーンを認識する段階、

ビデオ処理部と互いに連結するために台本の入力を受けて台本およびシーンを認識する段階、

認識されたビデオシーンおよび台本シーンを整列する前処理過程を実行する段階、

前処理されたビデオコンテンツを分析する段階、

前処理された台本コンテンツを分析する段階、

分析されたビデオコンテンツから生成されたアノテーションメタデータをアノテーション知識ベースに格納する段階、および

分析された台本コンテンツから生成されたナラティブメタデータをナラティブ知識ベースに格納する段階

を含み、

アノテーション知識ベースとナラティブ知識ベースは、エピソード、シーケンス、シーン、およびショットを共有することで相互補完的に使用してメタデータを自動で生成する、メタデータ自動生成方法であって、

ビデオ処理部と互いに連結するために台本の入力を受けて台本およびシーンを認識する段階は、

台本ショット抽出段階で、前処理段階で構造化された台本の階層構造と付加情報を入力としてシーン単位で分けられた台本と、ビデオとしてのショット、シーン情報とビデオの音声字幕の入力を受け、あるビデオ内の発話と台本と比較して整列して台本をショット単位で分け、

前処理された台本コンテンツを分析する段階は、

台本をビデオのようにショット単位で分けた後、登場人物、空間、オブジェクト分析により、ビデオ内に登場する登場人物、空間、オブジェクトをすべてそれぞれクラスタリング (Clustering) して登場する人物、空間、オブジェクトのカテゴリを把握し、登場人物の場合は、登場人物と登場人物があるショットとともに登場した回数をカウントして人物との関係とオブジェクトの名称を認識した後、登場人物、背景、空間をすべてセマンティック・ウェブとエンティティで連結させ、

ダイアログ分析段階では、すべての発話を分析して形態素 (POS)、個体名 (Named-Entity)、感情 (Emotion)、目的 (Intent)、言行 (Speech Act)、トーン (Tone)、敬称 (Honoriific) を認識し、すべての情報を各発話のメタデータとして格納し、

プロット分析段階で、分析された発話の言行と発話を分析して対話を事件単位で分け、

10

20

30

40

50

事件間の関係を分析して台本の起承転結を把握してナラティブメタデータとして格納する、メタデータ自動生成方法。

【請求項 6】

台本処理部と互いに連結するためにビデオの入力を受けてビデオのショットおよびシーンを認識する段階は、

人物およびオブジェクト認識により、あるショット内のショットの最初、最後、およびあるショット内でフレーム間の類似度が急変する 1 つ以上のフレームをキーフレームとして指定し、キーフレームイメージから人物の外形、顔とオブジェクトの外形、およびカテゴリを認識してメタデータとして格納する、

請求項 5 に記載のメタデータ自動生成方法。

10

【請求項 7】

前処理されたビデオコンテンツを分析する段階は、

人物とオブジェクトが認識されれば、フレームキャプショニング段階で、人物と人物、人物とオブジェクト、オブジェクトとオブジェクトのすべての相対位置を計算して場面を記述し、

場面が記述されれば、ショットキャプショニング段階で、人物のポーズと相対位置を利用して行動を認識し、

シーンキャプショニング段階で、認識された行動のパターンに基づいて人物の活動を区分してストーリー中心的構造に割り当てて格納し、多様な情報をビデオとして抽出してアノテーションメタデータとして格納する、

20

請求項 6 に記載のメタデータ自動生成方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、ビデオ (Video) とビデオの台本 (Script) からビデオコンテンツの多様な情報を記述し、ストーリーの理解をサポートするためにメタデータ (Metadata) を自動で生成する方法とそのシステムに関する。

【0002】

本発明は、科学技術情報通信部および情報通信技術振興センターの大学 ICT 研究センター育成支援事業の研究結果から導き出されたものである (課題管理番号: IITP-2017-0-01642、課題名: 人工知能を活用したコンテンツ創作技術)。

30

【背景技術】

【0003】

ビデオコンテンツのメタデータ (Metadata) は、ビデオ自体の情報を記述するメタデータと、ビデオと関連のある情報を記述するメタデータとに区分される。例えば、MPEG ビデオ規格の DASH 技術のうち、MPD (Media Presentation Description) のようにメディアの送信に適したデータモデルにビデオのセグメント (Segment) 情報を階層的に分離して表現するメタデータと、MPEG-4 LAsE R または MPEG-7 のようにビデオ、音声とともに、XML 形態でビデオコンテンツと関連する付加情報を記述するオブジェクトの位置、大きさ、登場時間、関連情報 URL などをアノテーション (Annotation) するメタデータとに分けられる。

40

【0004】

このようにビデオの付加情報を記述するメタデータを使用する代表的な例として、インタラクティブビデオ (Interactive Video) がある。インタラクティブビデオは、ビデオと相互作用が可能なオブジェクトなどを主にビデオ上に拡張し、ビデオ視聴者が該当のオブジェクトをクリックしたときに、ビデオと関連する付加情報を直ぐに取得することができて簡単に共有することができる新形態のビデオである。このようなインタラクティブビデオを生成するためには、視聴者の関心がありそうなビデオと関連する情報を選択し、これと関連する情報をインポートし、ビデオの特定の時点に多様な情報を

50

アノテーション (Annotation) してビデオのメタデータを生成する。

【0005】

しかし、このようなビデオコンテンツのメタデータを生成するためには多くの時間と費用がかかる。さらに、メタデータとして使用される、ビデオコンテンツと関連のある有意な情報は、ビデオだけで生成することが難しい。例えば、登場人物やビデオのストーリーに重要なオブジェクトがビデオから認識されたとしても、このような人物やオブジェクトの正確な名称を指定したり、人物と人物、事件と事件の関係性を類推したりする複雑な活用は決して簡単ではない。これを解決するために、ビデオからメタデータを生成することおよび台本を利用してメタデータを生成することで、ビデオの順序に合わせて多様な情報が盛り込まれたメタデータを自動で生成する方法が求められている。

10

【発明の概要】

【発明が解決しようとする課題】

【0006】

本発明が達成しようとする技術的課題は、ビデオコンテンツと関連する多様な情報を有するメタデータを自動で生成するために、ビデオから多様な情報を抽出する技術、台本から有意な情報を抽出する技術、およびビデオと台本との時間順序を整列してメタデータの使用性を高める方法とシステムを提供することにある。

【課題を解決するための手段】

【0007】

一側面において、本発明が提案する、ビデオコンテンツのメタデータ自動生成システムは、台本処理部と互いに連結するためにビデオの入力を受けてビデオのショットおよびシーンを認識して整列する前処理過程を実行し、ビデオコンテンツを分析するビデオ処理部、ビデオ処理部と互いに連結するために台本の入力を受けて台本およびシーンを認識して整列する前処理過程を実行し、台本コンテンツを分析する台本処理部、前記ビデオ処理部で生成されたアノテーションメタデータを格納するアノテーション知識ベース、および前記台本処理部で生成されたナラティブメタデータを格納するナラティブ知識ベースを含み、アノテーション知識ベースとナラティブ知識ベースは、エピソード、シーケンス、シーン、およびショットを共有することで相互補完的に使用してメタデータを自動生成する。

20

【0008】

ビデオ処理部は、入力されたビデオの毎フレーム間の類似度を測定し、類似度が予め定められた基準以上に急変するフレームの前後でビデオのショットを検出して区分し、区分されたビデオのすべてのショットで時空間が変わる時点を認識してビデオのシーンを検出する。

30

【0009】

台本処理部は、半構造化された台本書式を分析し、台本から人物、発話、説明文、場所、時間を区分するために多様な形態の台本構造様式と入力された台本構造とを比較して適した台本様式を選択し、構造化 (Structured) し、該当の様式が入力された台本によるものかを検証し、誤入力された様式が構造化結果として変形された場合は補正をし、台本の構造から台本のシーンを検出した情報をメタデータの階層構造として指定する。

【0010】

ビデオと台本のシーン順序を一致させるためにビデオのあるシーンの発話を音声認識技術によって字幕化して台本の発話と文字列比較アルゴリズムを利用して1次整列した後、ダイナミックプログラミング技法を利用してビデオと台本のシーン順序を整列する。

40

【0011】

台本をビデオのショット単位まで一致させるために、ビデオのシーン内に属するショット内の時間および音声テキスト情報を台本コンテンツ分析過程に伝達して処理する。

【0012】

また他の側面において、本発明が提案する、ビデオコンテンツのメタデータ自動生成方法は、台本処理部と互いに連結するためにビデオの入力を受けてビデオのショットおよびシーンを認識する段階、ビデオ処理部と互いに連結するために台本の入力を受けて台本お

50

よびシーンを認識する段階、認識されたビデオシーンおよび台本シーンを整列する前処理過程を実行する段階、前処理されたビデオコンテンツを分析する段階、前処理された台本コンテンツを分析する段階、分析されたビデオコンテンツから生成されたアノテーションメタデータを格納する段階、および分析された台本コンテンツから生成されたナラティブメタデータを格納する段階を含み、アノテーション知識ベースとナラティブ知識ベースは、エピソード、シーケンス、シーン、およびショットを共有することで相互補完的に使用してメタデータを自動生成する。

【0013】

台本処理部と互いに連結するためにビデオの入力を受けてビデオのショットおよびシーンを認識する段階は、人物およびオブジェクトの認識により、あるショット内でショットの最初、最後、およびあるショット内でフレーム間の類似度が急変する1つ以上のフレームをキーフレームとして指定し、キーフレームイメージから人物の外形、顔とオブジェクトの外形、およびカテゴリを認識してメタデータとして記録する。

10

【0014】

ビデオ処理部と互いに連結するために台本の入力を受けて台本およびシーンを認識する段階は、台本ショット抽出段階において、前処理段階で構造化された台本の階層構造と付加情報を入力としてシーン単位で分けられた台本と、ビデオとしてのショット、シーン情報とビデオの音声字幕を入力として受け、あるビデオ内の発話と台本とを比較して整列することによって台本をショット単位で分ける。

【0015】

20

前処理されたビデオコンテンツを分析する段階は、人物とオブジェクトが認識されれば、フレームキャプショニング段階において、人物と人物、人物とオブジェクト、オブジェクトとオブジェクトのすべての相対位置を計算し、場面を記述し、場面が記述されれば、ショットキャプショニング段階において、人物のポーズと相対位置を利用して行動を認識し、シーンキャプショニング段階で認識された行動のパターンに基づいて人物の活動を区分してストーリー中心的構造に割り当てて格納し、多様な情報をビデオで抽出してアノテーションメタデータとして格納する。

【0016】

前処理された台本コンテンツを分析する段階は、台本をビデオのようにショット単位で分けた後、登場人物、空間、オブジェクト分析により、ビデオ内に登場する登場人物、空間、オブジェクトをすべてそれぞれクラスタリング (Clustering) して登場する人物、空間、オブジェクトのカテゴリを把握し、登場人物の場合は、登場人物と登場人物があるショットにともに登場した回数をカウントして人物との関係、さらにオブジェクトの名称を認識した後、登場人物、背景、空間をすべてセマンティック・ウェブとエンティティで連結させ、ダイアログ分析段階ですべての発話を分析して形態素 (POS)、個体名 (Named-Entity)、感情 (Emotion)、目的 (Intent)、言行 (Speech Act)、トーン (Tone)、敬称 (Honoric) を認識、すべての情報を各発話のメタデータとして格納し、プロット分析段階で分析された発話の言行と発話を分析して対話を事件単位で分け、事件の関係を分析して台本の起承転結を把握してナラティブメタデータとして格納する。

30

【発明の効果】

【0017】

本発明の実施形態によると、ビデオコンテンツと関連する情報とビデオコンテンツの構造、さらにストーリーを理解することのできるビデオコンテンツの多様な情報を含んだメタデータを自動生成し、メタデータが活用される様々な場面で活用可能である。例えば、メタデータの活用は、生成されたメタデータで分けられたビデオ構造により、特定の時間に特定の情報をアノテーションするインタラクティブビデオのようにビデオコンテンツと関連する情報を簡単かつ迅速に得ようとするときにメタデータを活用することができ、ビデオコンテンツと関連する情報を検索 (Information Retrieval) する場合、ビデオコンテンツを要約して要約されたビデオを生成する場合、ビデオの登場

40

50

人物の発話特性を生かした人工知能チャットボット（Chatbot）を生成する場合、場面情報に基づいて対話のコンテキスト（Context）を認知するチャットボットを生成する場合のように、多様な用途に活用することができる。

【図面の簡単な説明】

【0018】

【図1】本発明の一実施形態における、ビデオコンテンツのメタデータ自動生成システムの構造を示した図である。

【図2】本発明の一実施形態における、ビデオコンテンツのメタデータ自動生成方法で前処理過程を説明するためのシステム構造を示した図である。

【図3】本発明の一実施形態における、ビデオコンテンツのメタデータ自動生成方法のうち、ビデオを分析してメタデータを生成する過程を示したフローチャートである。

10

【図4】本発明の一実施形態における、ビデオコンテンツのメタデータ自動生成方法のうち、台本を分析してメタデータを生成する過程を示したフローチャートである。

【図5】本発明の一実施形態における、ビデオコンテンツのメタデータ自動生成方法を説明するためのメタデータ蓄積過程を段階別に示したフローチャートである。

【図6】本発明の一実施形態における、生成したビデオコンテンツのメタデータを利用した活用例である。

【図7】本発明のまた他の実施形態における、生成したビデオコンテンツのメタデータを利用した活用例である。

【図8】本発明のさらに他の実施形態における、生成したビデオコンテンツのメタデータを利用した活用例である。

20

【発明を実施するための形態】

【0019】

以下、本発明の実施形態について、添付の図面を参照しながら詳しく説明する。

【0020】

図1は、本発明の一実施形態における、ビデオコンテンツのメタデータ自動生成方法を説明するためのシステム構造を簡略に示した図である。

【0021】

図1の自動ビデオコンテンツメタデータ生成システム100は、ビデオとビデオの台本の入力を受けてアノテーションメタデータとナラティブメタデータを出力するシステムであって、ビデオと台本をそれぞれ処理するビデオ処理部110と台本処理部130を含む。ビデオ処理部110と台本処理部130とを互いに連結するために、ビデオと台本のシーンを整列し、ショットを認識して整列する過程を前処理とし、台本とビデオからそれぞれ認識されたオブジェクトをシーンの順序に合うように整列する。アノテーションメタデータは、アノテーション知識ベース120に格納されて推論が可能であり、URIを通じて容易にメタデータにアクセス可能なオントロジベースのデータ格納空間である。ナラティブメタデータは、ナラティブ知識ベース130に格納され、アノテーション知識ベースと同じような役割を担う。アノテーション知識ベース120とナラティブ知識ベース130は、エピソード、シーケンス、シーン、およびショットなどを共有することで相互補完的に使用してメタデータを自動生成する。

30

【0022】

ビデオ処理部110は、入力されたビデオの毎フレーム間の類似度を測定し、類似度が予め定められた基準以上に急変するフレーム前後でビデオのショットを検出して区分し、区分されたビデオのすべてのショットで時空間が変わる時点を認識してビデオのシーンを検出する。

40

【0023】

台本処理部130は、半構造化された台本書式を分析し、台本から人物、発話、説明文、場所、時間を区分するために多様な形態の台本構造様式と入力された台本構造とを比較して適した台本様式を選択し、構造化（Structured）する。また、該当の様式が入力された台本によるものであるかを検証し、誤入力された様式が構造化結果によって

50

変形した場合は補正をし、台本の構造から台本のシーンを検出した情報をメタデータの階層構造として指定する。

【0024】

この後、ビデオと台本のシーンの順序を一致させるために、ビデオのあるシーンの発話を音声認識技術によって字幕化して台本の発話と文字列比較アルゴリズムを利用して1次整列した後、ダイナミックプログラミング技法を利用してビデオと台本のシーンの順序を整列する。このとき、台本をビデオのショット単位まで一致させるために、ビデオのシーン内に属するショット内の時間および音声テキスト情報を台本コンテンツ分析過程に伝達して処理する。

【0025】

図2は、本発明の一実施形態における、ビデオコンテンツのメタデータ自動生成方法で前処理過程を説明するためのシステム構造を示した図である。

【0026】

提案するビデオコンテンツのメタデータ自動生成方法は、台本処理部と互いに連結するためにビデオの入力を受けてビデオのショットおよびシーンを認識する段階、ビデオ処理部と互いに連結するために台本の入力を受けて台本およびシーンを認識する段階、認識されたビデオシーンおよび台本シーンを整列する前処理過程を実行する段階、前処理されたビデオコンテンツを分析する段階、前処理された台本コンテンツを分析する段階、分析されたビデオコンテンツから生成されたアノテーションメタデータを記録する段階、および分析された台本コンテンツから生成されたナラティブメタデータを記録する段階を含む。図2～4を参照しながら、ビデオコンテンツのメタデータ自動生成方法についてさらに詳しく説明する。

【0027】

図2を参照すると、ショットおよびシーン区分段階210、台本検証およびシーン検出段階230、ビデオシーンおよび台本シーン整列段階240は、本方法の前処理部分に該当し、ビデオと台本から多様な情報を抽出するビデオコンテンツ分析段階220と台本コンテンツ分析段階250前に、ビデオのショット、シーンを検出して台本のシーンを区分する段階である。まず、ショットおよびシーン区分段階210は、入力されたビデオの毎フレーム(Frame)間の類似度を測定し、類似度が急変するフレーム前後でビデオのショットを検出(Shot Boundary Detection)して区分する。このように区分されたビデオのすべてのショットで時空間が変わる時点を認識してビデオのシーンを検出する。台本検証およびシーン検出段階230では、半構造化された(Semi-Structured)台本書式を分析し、台本から人物、発話、説明文、場所、時間を区分するために多様な形態の台本構造様式と入力された台本構造とを比較して適した台本様式を選択し、簡単に処理できるように構造化(Structured)し、該当の様式が入力された台本によるものであるかを検証し、誤入力された様式、例えば、人物名ー発話の順に出現しなければならない台本が構造化結果では人物名ー人物名のように台本の構造が変形した場合には適切に補正をする。また、台本の構造から台本のシーンを検出し、その情報をメタデータの階層構造(Hierarchy)として指定する。このようにビデオのショット、シーン、台本のシーン情報を抽出した後、ビデオシーンおよび台本シーン整列段階240でビデオと台本のシーンの順序を整列する。シーンの順序を一致させるために、ビデオのあるシーンの発話を音声認識(Speech To Text)技術によって字幕化して台本の発話と文字列比較アルゴリズムを利用して1次整列した後、ダイナミックプログラミング(Dynamic Programming)技法を利用してビデオと台本のシーンの順序を整列する。このとき、ビデオのショットと台本の順序を一致させることは簡単ではない。その理由としては、台本はビデオの単位であるショット単位ではなくシーン単位で分けられているためであり、ショット単位まで一致させるために、ビデオのシーン内に属するショット内の時間および音声テキスト情報260を台本コンテンツ分析過程に伝達して処理する。

【0028】

10

20

30

40

50

図3は、本発明の一実施形態における、ビデオコンテンツのメタデータ自動生成方法のうち、ビデオを分析してメタデータを生成する過程を示したフローチャートである。

【0029】

先の前処理過程からビデオのショット、シーン階層構造270を通じて伝達されれば、まず、人物およびオブジェクト認識310により、あるショット内でショットの最初、最後、さらにあるショット内でフレーム（Frame）間の類似度が急変した1つ以上のフレームをキーフレームとして指定し、このキーフレームイメージから人物の外形（Contour）、顔、オブジェクトの外形、およびそのカテゴリを認識してメタデータとして格納する。人物とオブジェクトが認識されれば、フレームキャプショニング320段階において、人物の位置だけを認識するのではなく、人物と人物、人物とオブジェクト、オブジェクトとオブジェクトのすべての相対位置を計算し、人物の場合は人物の腕、脚、体の位置と形状を利用してポーズ（Pose）を認識して場面を記述する。このような情報が記述されれば、ショットキャプショニング330段階において、人物のポーズと相対位置を利用して行動（Action）を認識し、シーンキャプショニング340段階において、認識された行動のパターンに基づいて人物の活動（Activity）を区分し、それぞれ事件（Event）というストーリー中心的構造に割り当てて記録する。このように、多様な情報をビデオで抽出してアノテーションメタデータとして記録する。

【0030】

図4は、本発明の一実施形態における、ビデオコンテンツのメタデータ自動生成方法のうち、台本を分析してメタデータを生成する過程を示したフローチャートである。

【0031】

台本ショット抽出410段階では2つの入力を受けるようになるが、前処理段階で構造化された台本の階層構造と付加情報280を入力としてシーン単位で分けられた台本と、ビデオのショット、シーン情報260とビデオの音声字幕の入力を受けてあるビデオ内の発話と台本と比較、整列して台本をショット単位で分ける段階である。台本をビデオと同じようにショット単位で分けた後、登場人物、空間、オブジェクト分析420により、ビデオ内に登場する登場人物、空間、オブジェクトをすべてそれぞれクラスタリング（Clustering）して登場する人物、空間、オブジェクトのカテゴリを把握し、登場人物の場合は、登場人物と登場人物があるショットにともに登場した回数をカウントすることによって人物との関係とオブジェクトの名称を認識した後、登場人物、背景、空間をすべてセマンティック・ウェブとエンティティでEntity Linkingさせ、より多くの付加情報をインポートできるようにする。ダイアログ分析430段階では、すべての発話を分析して形態素（POS）、個体名（Named-Entity）、感情（Emotion）、目的（Intent）、言行（Speech Act）、トーン（Tone）、敬称（Honoriific）を認識し、すべての情報を各発話のメタデータとして記録する。次に、プロット分析440では、分析された発話の言行と発話を分析して対話を事件（Event）単位で分け、多くの事件間の関係を分析して台本の起承転結を把握してメタデータとして記録する。このように生成されたメタデータは、ナラティブメタデータとして格納する。

【0032】

図5は、本発明の一実施形態における、ビデオコンテンツのメタデータ自動生成方法を説明するためのメタデータ蓄積過程を段階別に示したフローチャートである。

【0033】

まず、ビデオと台本の入力510を受け、台本を構造化520してビデオと台本の階層構造を分析してシーン、ショットの情報を追加する。この後、ビデオと台本内から多様な情報を認識して抽出してコンテンツ530情報を追加する。次に、ビデオと台本のストーリーを分析するために行動および発話の特性を分析し、事件540情報を生成して追加し、最後にメタデータ550を生成して知識ベースに追加する。

【0034】

図6は、本発明の一実施形態における、生成したビデオコンテンツのメタデータを利用

した活用例である。

【0035】

ナラティブメタデータの情報を利用してビデオコンテンツのストーリーから重要な場面610を抽出し、サムネイルのスライドバー形態で下側に表示する。ユーザは、関心のある場面を直接クリックして場面を再生させることもできるし、画面上の検索ウィンドウ620を利用して直接検索することもできる。検索は、予約語ではない自然語(Natural Language)で入力すれば、形態素分析、個体名認識により、登場人物、場所、行動のような主要単語を自動で抽出し、SPARQL Queryを自動で生成し、ナラティブ知識ベースとアノテーション知識ベースのSPARQL Endpointによって知識を検索し、該当する場面のビデオをインポートして画面に表示する。ビデオコンテンツと関連する情報を相互作用が可能な拡張オブジェクトとしてアノテーションし、認識された人物やオブジェクトの外郭線と人物の名称630を画面上に記すことによってクリック可能なオブジェクトであることを表示する。人物をクリックする場合は、人物の付加情報と人物と人物の関係図、人物と関連するシーン、事件などの情報が表示され、オブジェクトをクリックする場合は、オブジェクトがストーリーでどのような意味をもつものであるかを表示し、オブジェクトがインターネットで購入可能なものであれば、購入するための追加オブジェクト650を表示して直ぐに購入できるようにすることも可能である。

10

【0036】

図7は、本発明のまた他の実施形態における、生成したビデオコンテンツのメタデータを利用した活用例である。

20

【0037】

N-Screen環境において、TVスクリーン710は、ビデオ再生プログラムとしてアノテーションメタデータとナラティブメタデータを生成するために使用したビデオを再生する。ビデオが再生されるときにモバイルデバイス720でも同時に同じビデオを再生し、TVスクリーンのビデオとモバイルデバイスのビデオをソケット通信によってリアルタイムで再生時間を合わせて同期化730する。このように同期化されたビデオは、モバイルデバイスにビデオのように示されるビデオと関連するストーリー情報を表示するときにも、ビデオ時間と同期化して表示する。ビデオ上の顔バウンディングボックス740は、ビデオに登場する人物のうちで該当の場面のストーリーで重要な人物の顔をアノテーションメタデータのデータを利用して表示する。ナラティブメタデータにあるビデオのストーリー情報を要約し、人物関係750と場面ストーリー要約760のようなビデオと関連する付加情報を表示する。人物関係は、該当の場面のストーリーで極めて重要な人物の人物関係図を表示し、ビデオのストーリーを簡単に把握して登場人物の関係を把握するために表示する。場面ストーリーの要約は、該当の場面をナラティブメタデータのデータを利用して要約して表示する。このような付加情報により、ビデオを見ながらユーザが把握することのできなかつた追加情報を、TVで再生中のビデオを止めてインターネットで検索する必要がなく、モバイルデバイスで直観的に表示することができ、ストーリーを簡単に理解できるようにする。

30

【0038】

図8は、本発明のさらに他の実施形態における、生成したビデオコンテンツのメタデータを利用した活用例である。

40

【0039】

アノテーションメタデータを生成するとき、キーフレームイメージ上のオブジェクトを認識し、オブジェクトの登場時間とオブジェクトの位置を認識してメタデータに記録する。ビデオとビデオ内のオブジェクトと関連する情報をインポートしてビデオ上に情報を拡張するために、メタデータに記録されたオブジェクトの登場時間と位置、大きさ、オブジェクトの情報を読み込む。また、キーフレーム間のフレーム上にオブジェクトと関連する情報をアノテーションするために補間(interpolation)を行って中間値である特定の時間のオブジェクト位置と大きさを生成し、すべてのフレームにアノテーシ

50

ョンする。オブジェクトの情報は、認識されたオブジェクトのビデオとオブジェクトのイメージとして類似するショッピングモールの製品広告をインポートして示したり、ナラティブメタデータの現在の場面と関連するオブジェクト、例えば、プレゼントのようなオブジェクトの情報をインポートして示したりする。また、製品広告は、ユーザに応じて異なる情報を推薦して表示する。

【0040】

上述した装置は、ハードウェア構成要素、ソフトウェア構成要素、および／またはハードウェア構成要素とソフトウェア構成要素との組み合わせによって実現されてよい。例えば、実施形態で説明された装置および構成要素は、例えば、プロセッサ、コントローラ、ALU (arithmetic logic unit)、デジタル信号プロセッサ、マイクロコンピュータ、FPA (field programmable array)、PLU (programmable logic unit)、マイクロプロセッサ、または命令を実行して応答することができる様々な装置のように、1つ以上の汎用コンピュータまたは特殊目的コンピュータを利用して実現されてよい。処理装置は、オペレーティングシステム(OS)およびOS上で実行される1つ以上のソフトウェアアプリケーションを実行してよい。また、処理装置は、ソフトウェアの実行にตอบสนองし、データにアクセスし、データを記録、操作、処理、および生成してもよい。理解の便宜のために、1つの処理装置が使用されるとして説明される場合もあるが、当業者は、処理装置が複数の処理要素および／または複数種類の処理要素を含んでもよいことが理解できるであろう。例えば、処理装置は、複数のプロセッサまたは1つのプロセッサおよび1つのコントローラを含んでよい。また、並列プロセッサのような、他の処理構成も可能である。

【0041】

ソフトウェアは、コンピュータプログラム、コード、命令、またはこれらのうちの1つ以上の組み合わせを含んでもよく、思うままに動作するように処理装置を構成したり、独立的または集散的に処理装置に命令したりしてよい。ソフトウェアおよび／またはデータは、処理装置に基づいて解釈されたり、処理装置に命令またはデータを提供したりするために、いかなる種類の機械、コンポーネント、物理装置、仮想装置、コンピュータ記録媒体または装置に具現化されてよい。ソフトウェアは、ネットワークによって接続されたコンピュータシステム上に分散され、分散された状態で記録されても実行されてもよい。ソフトウェアおよびデータは、1つ以上のコンピュータ読み取り可能な記録媒体に記録されてよい。

【0042】

実施形態に係る方法は、多様なコンピュータ手段によって実行可能なプログラム命令の形態で実現されてコンピュータ読み取り可能な媒体に記録されてよい。前記コンピュータ読み取り可能な媒体は、プログラム命令、データファイル、データ構造などを単独または組み合わせて含んでよい。前記媒体に記録されるプログラム命令は、実施形態のために特別に設計されて構成されたものであっても、コンピュータソフトウェアの当業者に公知な使用可能なものであってもよい。コンピュータ読み取り可能な記録媒体の例としては、ハードディスク、フロッピディスク、磁気テープのような磁気媒体、CD-ROM、DVDのような光媒体、フロッティカルディスク (floptical disk) のような光磁気媒体、およびROM、RAM、フラッシュメモリなどのようなプログラム命令を記録して実行するように特別に構成されたハードウェア装置が含まれる。プログラム命令の例は、コンパイラによって生成されるもののような機械語コードだけではなく、インタプリタなどを使用してコンピュータによって実行される高級言語コードを含む。

【0043】

以上のように、実施形態を、限定された実施形態および図面に基づいて説明したが、当業者であれば、上述した記載から多様な修正および変形が可能であろう。例えば、説明された技術が、説明された方法とは異なる順序で実行されたり、かつ／あるいは、説明されたシステム、構造、装置、回路などの構成要素が、説明された方法とは異なる形態で結合されたりまたは組み合わせられたり、他の構成要素または均等物によって対置されたり置換

されたとしても、適切な結果を達成することができる。

【 0 0 4 4 】

したがって、異なる実施形態であっても、特許請求の範囲と均等なものであれば、添付される特許請求の範囲に属する。

10

20

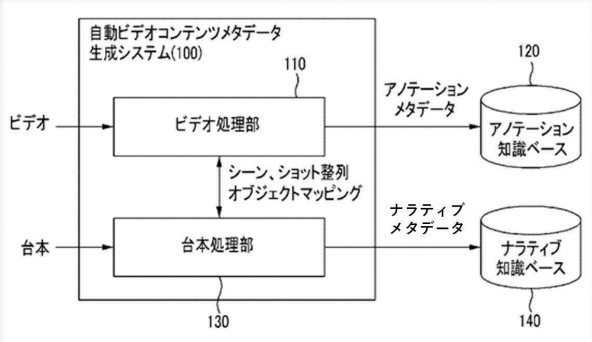
30

40

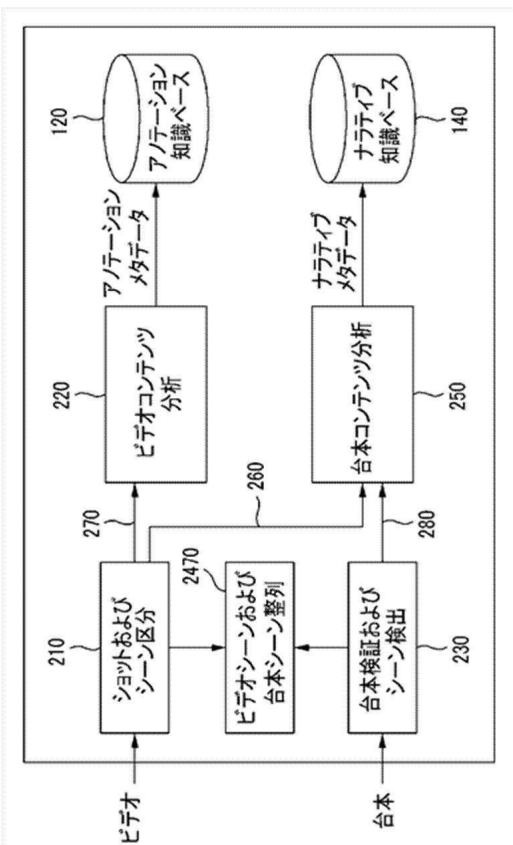
50

【図面】

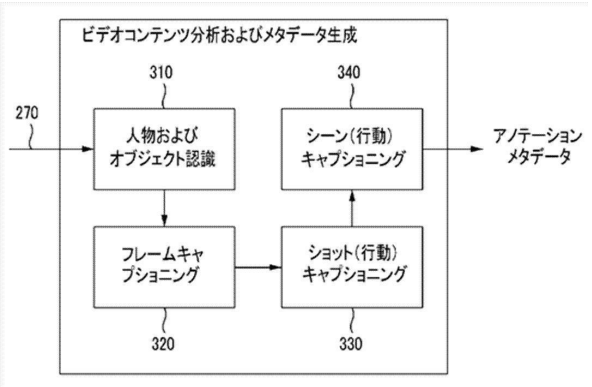
【図 1】



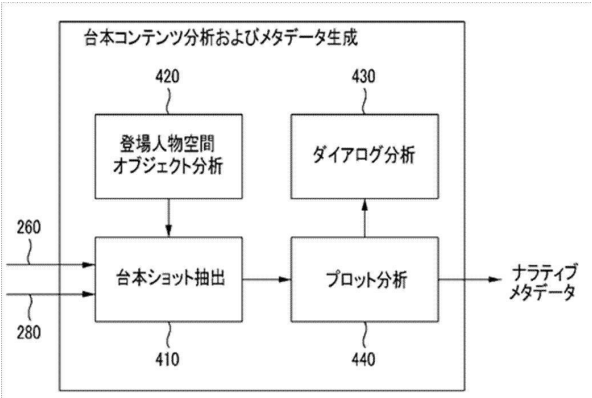
【図 2】



【図 3】



【図 4】



10

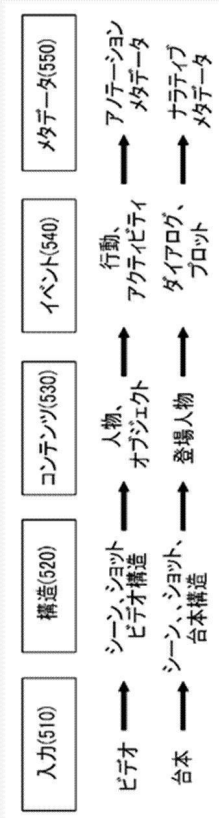
20

30

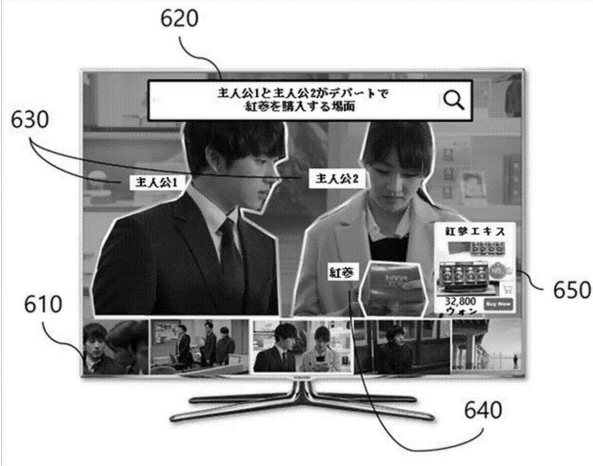
40

50

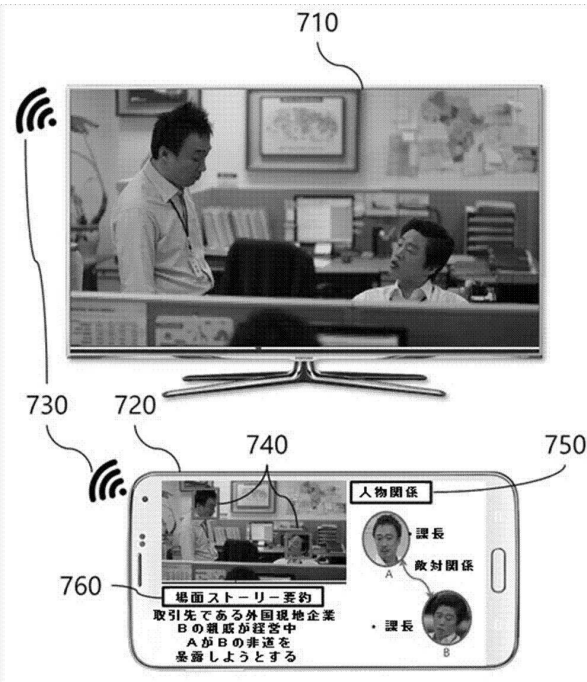
【図 5】



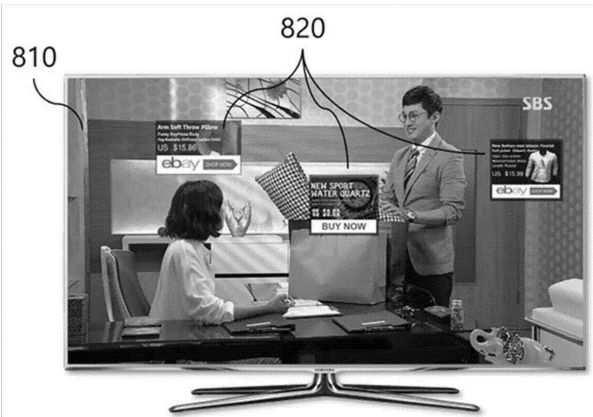
【図 6】



【図 7】



【図 8】



10

20

30

40

50

フロントページの続き

- (56)参考文献 米国特許出願公開第2011/0087703 (US, A1)
特開2003-009069 (JP, A)
米国特許出願公開第2018/0213289 (US, A1)
桑野 秀豪, 外2名, 映像・音声認識, 言語処理の適用による経済化メタデータ生成技術, N
TT技術ジャーナル, 日本, 社団法人電気通信協会, 2004年05月01日, 第16巻, 第5号, p
p. 22-25
- (58)調査した分野 (Int.Cl., DB名)
H04N 21/00-21/858