

明 細 書

発明の名称：

監視装置、監視システム、監視方法、及びプログラムが格納された非一時的なコンピュータ可読媒体

技術分野

[0001] 本開示は、監視装置、監視システム、監視方法、及びプログラムが格納された非一時的なコンピュータ可読媒体に関する。

背景技術

[0002] 近年、1人で営業している店舗などを狙った強盗などの犯罪が増えている。これを防ぐために、監視カメラを設置して警備会社に監視を委託している店舗なども増えている。しかし、多数の顧客と契約している警備会社は、個々の監視カメラ映像を常に見ているわけではなく、非常ボタンなどで通報しなければ見て貰えないのが現実である。加えて、例えば強盗から脅されているなどの理由で、従業員が非常ボタンを押せない場合もある。このため、監視カメラの映像を人が監視する監視方法には限界がある。

[0003] この問題を解決するために、監視カメラの映像をコンピュータで監視する、インテリジェントな監視方法が提案されている。例えば、特許文献1には、監視カメラだけでなくマイクも設置し、取得した映像や音などをプログラムで解析して異常事態の発生を検知する、監視方法が開示されている。

[0004] 一般的に、映像から異常を検知する場合、特許文献1に記されている様に、監視カメラの映像データをネットワーク経由で収集し、コンピュータにて分析を行う。映像分析では、特定人の顔画像や単一または複数人の異常行動や特定の場所における放置物など、危険に繋がる映像特徴を事前に登録しておき、その特徴の存在を検知している。

[0005] また、映像だけでなく、特許文献1の様に、音による異常検知も行われている。音は、人の発話内容を認識して分析する音声認識と、音声以外の音を分析する音響分析が存在するが、そのどちらとも、さほどのコンピュータ資

源は必要としない。このため、例えばスマートフォンに搭載されている様な組み込み向けのCPU（Central Processing Unit）であっても、リアルタイムに分析することが十分に可能である。

[0006] 音を分析することによる異常事態の発生を検知は、不測の異常事態に対しても有効である。なぜならば、異常事態に遭遇した人が悲鳴や大声を出すことは、普遍的な自然法則であるからである。

[0007] また、音は360度の全ての方向に拡散し、暗闇の中でも伝搬し、途中に障害物があっても回り込む性質を持っている。このため、音による監視の場合、カメラの様に視野や方向や照明によって監視対象が制限されることはなく、暗闇や物陰で発生した悲鳴や大声でも見逃さないという、監視に適した優れた特徴を有している。

[0008] さらに、複数のマイクにて音を収集する場合、特許文献2に開示されている様に、音源から各々のマイクまでの音の到達時刻差や、音の拡散及び減衰による音圧差などから、音源の位置を推定することができる。

[0009] また、特許文献3には、画像に映った人物の関節位置から、姿勢を推定する技術が開示されている。これを映像に適用することで、腕や手の動きなどから、人物の行動が推定される。

[0010] また、特許文献4には、人間の顔画像から表情を認識する、表情認識という技術が開示されている。

先行技術文献

特許文献

[0011] 特許文献1：特開2013-131153号公報

特許文献2：特表2013-545382号公報

特許文献3：特開2021-086322号公報

特許文献4：国際公開第2019/102619号

発明の概要

発明が解決しようとする課題

[0012] 映像から得られる特徴だけに着目して、不測の異常事態の発生を検知することは難しい。すなわち、映像特徴単独で異常事態の発生を検知することができるような映像特徴を事前に定義するのは難しい。映像を分析して、異常の発生を検知する場合、それぞれの異常に対応する映像の特徴を予め定義しておく必要がある。すなわち、映像から異常事態の発生を検知するには、様々な異常事態に対して映像特徴を事前に定義した上で、分析のためのプログラム（例えば、機械学習により分類器を生成するプログラムなど）を用意しなければならない。しかし、実社会では、犯罪被疑者や被害者の人相特徴、所持品、行動は多様で、様々な犯罪や事故が発生する。このため、何らかの前提条件が加わらない限り、異常事態に対応する映像特徴を事前に定義するのは困難で、映像だけから異常事態の発生を検知する方法は実用性に欠ける。

[0013] 例えば、特許文献1では、特定人の顔画像を事前に登録する事が例示されているが、不測の異常事態を引き起こす全ての人物の顔画像が事前に収集されているわけではないため、顔画像や人相特徴を映像特徴とする異常検知は用途が限られる。また、特許文献1には、単一または複数人の異常行動を事前に登録する事も例示されているが、例えばお客から代金を受け取ってレジから釣銭を渡す行動と、強盗から脅されてレジから現金を渡す行動との差異は少ない。このため、当事者の映像特徴から異常行動を判定することは難しい。

[0014] 一方で、上述の通り、音を分析することによる異常事態の発生の検知は、不測の異常事態に対しても有効である。しかしながら、音による分析だけでは、検知した異常事態に対して対応が必要であるか否かなどの評価をすることはできない。

[0015] そこで、本明細書に開示される実施形態が達成しようとする目的の1つは、異常事態の発生を検知し、異常事態を適切に把握することができる新規な技術を提供することである。

課題を解決するための手段

[0016] 本開示の第 1 の態様にかかる監視装置は、
監視対象エリアにおける異常事態の発生により人が発声した所定の音声を
取得する音声取得手段と、
前記所定の音声から得られる特徴に基づいて、前記所定の音声を発声した
人物を特定する人物特定手段と、
前記監視対象エリアを撮影するカメラの映像から、特定された前記人物を
検索し、当該人物の表情又は動作を分析する分析手段と、
前記分析の結果に基づいて、前記監視対象エリアにおける前記異常事態を
評価する異常事態評価手段と
を有する。

[0017] 本開示の第 2 の態様にかかる監視システムは、
監視対象エリアを撮影するカメラと、
監視対象エリアにおいて発生する音を検知するセンサと、
監視装置と
を備え、
前記監視装置は、
前記監視対象エリアにおける異常事態の発生により人が発声した所定の音
声を前記センサから取得する音声取得手段と、
前記所定の音声から得られる特徴に基づいて、前記所定の音声を発声した
人物を特定する人物特定手段と、
前記カメラの映像から、特定された前記人物を検索し、当該人物の表情又
は動作を分析する分析手段と、
前記分析の結果に基づいて、前記監視対象エリアにおける前記異常事態を
評価する異常事態評価手段と
を有する。

[0018] 本開示の第 3 の態様にかかる監視方法では、
監視対象エリアにおける異常事態の発生により人が発声した所定の音声を
取得し、

前記所定の音声から得られる特徴に基づいて、前記所定の音声を発声した人物を特定し、

前記監視対象エリアを撮影するカメラの映像から、特定された前記人物を検索し、当該人物の表情又は動作を分析し、

前記分析の結果に基づいて、前記監視対象エリアにおける前記異常事態を評価する。

[0019] 本開示の第4の態様にかかるプログラムは、

監視対象エリアにおける異常事態の発生により人が発声した所定の音声を取得する音声取得ステップと、

前記所定の音声から得られる特徴に基づいて、前記所定の音声を発声した人物を特定する人物特定ステップと、

前記監視対象エリアを撮影するカメラの映像から、特定された前記人物を検索し、当該人物の表情又は動作を分析する分析ステップと、

前記分析の結果に基づいて、前記監視対象エリアにおける前記異常事態を評価する異常事態評価ステップと

をコンピュータに実行させる。

発明の効果

[0020] 本開示によれば、異常事態の発生を検知し、異常事態を適切に把握することができる新規な技術を提供することができる。

図面の簡単な説明

[0021] [図1]実施の形態の概要にかかる監視装置の構成の一例を示すブロック図である。

[図2]実施の形態の概要にかかる監視装置の動作の流れの一例を示すフローチャートである。

[図3]実施の形態にかかる監視システムの構成の一例を示す模式図である。

[図4]音響センサの機能構成の一例を示すブロック図である。

[図5]分析サーバーの機能構成の一例を示すブロック図である。

[図6]コンピュータのハードウェア構成の一例を示す模式図である。

[図7]実施の形態にかかる監視システムの動作の流れの一例を示すフローチャートである。

[図8]図7に示したフローチャートにおけるステップS107における処理の流れの一例を示すフローチャートである。

発明を実施するための形態

[0022] <実施の形態の概要>

実施の形態の詳細を説明する前に、実施の形態の概要について説明する。

図1は、実施の形態の概要にかかる監視装置1の構成の一例を示すブロック図である。図1に示すように、監視装置1は、音声取得部2、人物特定部3、分析部4、及び異常事態評価部5を有し、所定の監視対象エリアを監視するための装置である。

[0023] 音声取得部2は、監視対象エリアにおける異常事態の発生により人が発声した所定の音声を取得する。ここで、所定の音声とは、人が異常事態に遭遇した場合に発声する音声であり、例えば、悲鳴や叫び声である。音声取得部2は、例えば、監視対象エリアに設置されたマイクにより集音された悲鳴や叫び声を取得する。

[0024] 人物特定部3は、音声取得部2が取得した所定の音声から得られる特徴に基づいて、所定の音声を発声した人物を特定する。例えば、人物特定部3は、所定の音声から得られる特徴と、予め登録された人物の音声の特徴とに基づいて、所定の音声を発声した人物が、音声の特徴が登録された人物のうちいずれの人物に該当するかを特定する。

[0025] 分析部4は、監視対象エリアを撮影するカメラの映像から、人物特定部3によって特定された人物を検索し、当該人物の表情又は動作を分析する。例えば、分析部4は、映像内の検索された人物の表情が、所定の表情であるか否かについて分析する。ここで、この所定の表情は、人が異常事態に遭遇した場合に表れる表情であり、具体的には例えば、怯えた表情や怒りの表情などである。また、例えば、分析部4は、映像内の検索された人物の動作が、所定の動作であるか否かについて分析する。ここで、この所定の動作は、例

えば、異常事態に遭遇した人物が行う一連の動作であってもよいし、ジェスチャーであってもよい。なお、分析部４は、表情の分析と動作の分析のうち、いずれか一方だけを行ってもよいし、両方を行ってもよい。

[0026] 異常事態評価部５は、分析部４の分析の結果に基づいて、監視対象エリアにおける異常事態を評価する。例えば、異常事態評価部５は、対応が必要な異常事態であるか否かを判定するための指標（例えば、スコア）を算出する。また、異常事態評価部５は、その指標に基づいて、対応が必要な異常事態であるか否かを判定してもよい。

[0027] 図２は、実施の形態の概要にかかる監視装置１の動作の流れの一例を示すフローチャートである。以下、図２に沿って、監視装置１の動作の流れの一例について説明する。

[0028] まず、ステップＳ１１において、音声取得部２が、監視対象エリアにおける異常事態の発生により人が発声した所定の音声を取得する。

次に、ステップＳ１２において、人物特定部３が、音声取得部２が取得した所定の音声から得られる特徴に基づいて、所定の音声を発声した人物を特定する。

次に、ステップＳ１３において、分析部４が、監視対象エリアを撮影するカメラの映像から、人物特定部３によって特定された人物を検索し、当該人物の表情又は動作を分析する。

次に、ステップＳ１４において、異常事態評価部５は、分析部４の分析の結果に基づいて、監視対象エリアにおける異常事態を評価する。

[0029] 以上、実施の形態の概要にかかる監視装置１について説明した。監視装置１によれば、音声及び映像を用いた処理が行われ、これにより、異常事態の発生を検知し、異常事態を適切に把握することができる。

[0030] ＜実施の形態の詳細＞

次に、実施の形態の詳細について説明する。

図３は、実施の形態にかかる監視システム１０の構成の一例を示す模式図である。本実施の形態では、監視システム１０は、分析サーバー１００と、

監視カメラ２００と、音響センサ３００とを備えている。監視システム１０は、所定の監視対象エリア９０を監視するためのシステムである。監視対象エリア９０は、例えば、店舗や金融機関などであるが、これに限らず、監視が行われる任意のエリアであってもよい。

[0031] 監視カメラ２００は、監視対象エリア９０を撮影するために設置されたカメラである。監視カメラ２００は、監視対象エリア９０を撮影し、映像データを生成する。監視対象エリア９０の全体を監視できる適切な位置に監視カメラ２００が設置されている。なお、監視対象エリア９０の全体を監視するために、複数の監視カメラ２００が設置されてもよい。

[0032] 本実施の形態では、音響センサ３００は、監視対象エリア９０内の各所に設けられている。具体的には、例えば、１０メートルから２０メートル程度の間隔で、音響センサ３００が設置されている。音響センサ３００は、監視対象エリア９０の音を集音して分析する。具体的には、音響センサ３００は、マイクとサウンドデバイスとＣＰＵなどから構成され、音をセンシングする機器である。音響センサ３００は、周囲の音をマイクで集音してサウンドデバイスにてデジタル信号へ変換し、その後、ＣＰＵにより音響分析を行う。この音響分析では、例えば、悲鳴や叫び声などの異常音の検知が行われる。なお、この音響センサ３００には、音声認識の機能を搭載してもよい。その場合、叫び声などの発話内容を認識して、異常事態の深刻度を推定するなど、より高度な分析が可能になる。

[0033] 本実施の形態において、音響センサ３００を１０メートルから２０メートル程度の間隔で監視対象エリア９０内の各所に設置したのは、エリア内の何処で異常音が発生しても、複数の音響センサ３００が検知できるようにするためである。一般に、店舗などでの騒音は６０デシベル程度であるのに対し、悲鳴や叫び声は８０から１００デシベル程度の大きさを有する。しかし、例えば音の発生位置から１０メートル離れた場合、音源付近では１００デシベルだった異常音も８０デシベルまで減衰する。ここで、音源から音響センサ３００までの距離が離れすぎると、音響センサ３００の位置にて６０デシ

ベル程度の背景の騒音と、減衰した異常音を識別するのが困難になる。このため、本実施の形態では、上述のような間隔で音響センサ３００を配置している。なお、どこまで間隔を離しても、複数の音響センサ３００が同一の異常音を検知できるかは、背景の騒音レベルや各音響センサ３００の性能に依存するため、必ずしも１０メートルから２０メートルという配置の制約があるわけではない。

[0034] 分析サーバー１００は、監視カメラ２００及び音響センサ３００によって得られたデータを分析するためのサーバーであり、図１に示した監視装置１の機能を備えている。分析サーバー１００は、音響センサ３００から分析結果を受け取ると共に、必要に応じて監視カメラ２００から映像データを取得して映像の分析を行う。分析サーバー１００と、監視カメラ２００はネットワーク５００で通信可能に接続されている。同様に、分析サーバー１００と、音響センサ３００はネットワーク５００で通信可能に接続されている。ネットワーク５００は、監視カメラ２００と音響センサ３００と分析サーバー１００との間の通信を伝送するネットワークであり、有線ネットワークでもよいし、無線ネットワークでもよい。

[0035] 図４は、音響センサ３００の機能構成の一例を示すブロック図である。また、図５は、分析サーバー１００の機能構成の一例を示すブロック図である。

[0036] 図４に示すように、音響センサ３００は、異常検知部３０１と一次判定部３０２とを有する。

[0037] 異常検知部３０１は、音響センサ３００が検知した音に基づいて、監視対象エリア９０内での異常事態の発生を検知する。異常検知部３０１は、例えば、音響センサ３００が検知した音が、所定の異常音（具体的には、例えば悲鳴や叫び声などの音声）に該当するか否かを判定することにより、異常事態の発生を検知する。すなわち、異常検知部３０１は、音響センサ３００が検知した音が所定の異常音に該当する場合、監視対象エリア９０内で異常事態が発生したと判断する。また、本実施の形態では、異常検知部３０１は、

異常事態が発生したと判断した場合、異常度合いを示すスコアを算出する。
例えば、異常検知部３０１は、音声の大きさが大きいほど高いスコアを算出してもよい。

[0038] 一次判定部３０２は、異常事態の発生が検知された場合に、この異常事態に対して対応が不要であるか否かを判定する。例えば、一次判定部３０２は、異常検知部３０１が算出したスコアと、事前に設定した閾値とを比較することにより、この判定を行う。すなわち、一次判定部３０２は、算出されたスコアが閾値以下である場合、検知された異常事態に対して対応が不要であると判定する。この場合、監視システム１０における更なる処理は行われなない。これに対し、異常事態に対して対応が不要であると判定されなかった場合、音響センサ３００から分析サーバー１００に対し異常事態の発生が通知される。なお、この通知処理は、異常検知部３０１の処理として行われてもよい。

[0039] 音響センサ３００から分析サーバー１００に対し異常事態の発生が通知されると、後述する分析サーバー１００の処理が行われる。このように、本実施の形態では、一次判定部３０２の判定結果に応じて分析サーバー１００の処理が行われるか否かが決まるが、一次判定部３０２の判定結果によらずに分析サーバー１００の処理が行われてもよい。すなわち、異常検知部３０１が異常事態の発生を検知した全ての場合において、分析サーバー１００の処理が行われてもよい。つまり、一次判定部３０２による判定処理は省略されてもよい。

[0040] 図５に示すように、分析サーバー１００は、音声取得部１０１、人物特定部１０２、音源位置推定部１０３、映像取得部１０４、人物検索部１０５、表情認識部１０６、動作認識部１０７、表情スコア算出部１０８、動作スコア算出部１０９、二次判定部１１０、信号出力部１１１、音声特徴記憶部１２１、外観特徴記憶部１２２、異常行動記憶部１２３、及びジェスチャー記憶部１２４を有する。

[0041] 音声取得部１０１は、監視対象エリア９０における異常事態の発生により

人が発声した所定の音声を取得する。具体的には、音響センサ３００によって検知された所定の音声（悲鳴や叫び声）を音響センサ３００から取得する。具体的には、音声取得部１０１は、音響センサ３００から分析サーバー１００に異常事態の発生が通知されると、音響センサ３００から音声を取得する。

[0042] 人物特定部１０２は、音声取得部１０１が取得した所定の音声から得られる特徴に基づいて、この所定の音声を発声した人物を特定する。本実施の形態では、人物特定部１０２は、音声特徴記憶部１２１に記憶された音声の特徴と、音声取得部１０１が取得した所定の音声から得られる特徴とを照合することにより、この所定の音声を発声した人物を特定する。

[0043] 音声特徴記憶部１２１は、監視対象エリア９０に存在しうる人物（例えば、従業員など）毎に、人物の識別情報と当該人物の音声の特徴とを対応付けて記憶するデータベースである。人物特定部１０２は、音声の特徴を比較することによって、所定の音声を発声した人物が、音声の特徴が登録された人物のうちいずれの人物に該当するかを特定する。なお、音声の特徴としては、例えば、フォルマントの基底周波数や声帯の開閉に伴うゆらぎなどが用いられるが、これらに限られない。人物特定部１０２は、人物を特定するために、音声取得部１０１が取得した音声から、所定の音声解析処理を行って、特徴を抽出する。

[0044] なお、人物特定部１０２は、音声取得部１０１が取得した所定の音声に対応する人物として、必ずしも１人の人物を特定しなくてもよい。人物特定部１０２は、音声取得部１０１が複数の人物の音声を取得した場合、それら複数の人物をそれぞれ特定してもよい。また、音声取得部１０１が取得した１人の音声に対して、必ずしも１人の人物を特定しなくてもよい。例えば、類似の音声の特徴を有する人物が複数人、登録されていた場合には、人物特定部１０２は、所定の音声を発声した複数人の候補を特定してもよい。

[0045] 音源位置推定部１０３は、監視対象エリア９０に設けられた音響センサ３００が検知した音の発生源を推定することにより、異常事態の発生位置を推

定する。具体的には、音源位置推定部１０３は、複数の音響センサ３００から分析サーバー１００に異常事態の発生が通知されると、これら複数の音響センサ３００から収集した音声データに対して、例えば特許文献２などに開示されている公知の音源位置推定処理を行なう。すなわち、例えば、音源位置推定部１０３は、監視対象エリア９０の複数の位置にそれぞれ設けられたマイクへの音声の到達時刻差や、音の拡散及び減衰による音圧差などから、当該音声の音源位置を推定してもよい。これにより、音源位置推定部１０３は、所定の音声の音源位置、すなわち異常事態の発生位置を推定する。

[0046] 映像取得部１０４は、音源位置推定部１０３により異常事態の発生位置が推定されると、推定された位置を撮影している監視カメラ２００から、映像データを取得する。なお、例えば、分析サーバー１００には各監視カメラ２００がどのエリアを撮影しているかを表す情報が予め記憶されており、映像取得部１０４はこの情報と推定された位置とを比較することにより、推定された位置を撮影している監視カメラ２００を特定する。

[0047] 人物検索部１０５は、異常事態の発生位置の近傍の映像から、異常音を発した人物を検索する。すなわち、人物検索部１０５は、映像取得部１０４が取得した映像から、人物特定部１０２により特定された人物を検索する。なお、人物特定部１０２が複数人の人物を特定した場合には、人物検索部１０５は、これら複数人の人物に対して検索処理を行う。本実施の形態では、人物検索部１０５は、外観特徴記憶部１２２に記憶された人物の外観の特徴と、映像取得部１０４が取得した映像から抽出される人物の外観の特徴とを照合することにより、映像の中から人物特定部１０２により特定された人物を検索する。

[0048] 外観特徴記憶部１２２は、監視対象エリア９０に存在しうる人物（例えば、従業員など）毎に、すなわち、音声特徴記憶部１２１に音声の特徴が登録された人物毎に、人物の識別情報と当該人物の外観の特徴とを対応付けて記憶するデータベースである。人物検索部１０５は、具体的には、映像から人物を検出し、各々の人物の外観の特徴を抽出し、外観特徴記憶部１２２に事前に

登録された人物の外観の特徴と照合することによって、人物特定部102により特定された人物を検索する。ここで、外観の特徴としては、顔の特徴であってもよいし、服装や帽子などの特徴であってもよいし、従業員が装着するIDカードに印刷されたコード（バーコード又は二次元コードなど）であってもよい。すなわち、外観の特徴は、映像から取得することができる、人物毎に異なる外観の特徴であればよい。人物検索部105は、人物を検索するために、映像取得部104が取得した映像から、所定の画像解析処理を行って、特徴を抽出する。人物検索部105は、人物を検索すると、検索された人物を映像内で特定するアノテーションを映像データに付与する。

[0049] 表情認識部106は、人物特定部102によって特定された人物、すなわち人物検索部105によって検索された人物の表情を認識する。具体的には、例えば、表情認識部106は、上述したアノテーションが付与された映像データに対して公知の表情認識処理を行って表情（平穏、笑い、怒り、怯えなどの心理状態を表す表情）を認識する。例えば、表情認識部106は、顔画像に対して特許文献4で開示された処理を行うことで表情を認識してもよい。特に、表情認識部106は、人物の顔に表れた表情が所定の表情であるか否かを分析する。ここで、所定の表情は、具体的には例えば、怯えた表情や怒りの表情などである。例えば異常な音声を発した人物が店員だったとする。この場合、通常であれば笑顔で接客するはずの店員が、強盗などの異常事態に遭遇すると、笑顔が失せて、恐怖に怯えるなどの表情に変化する。したがって、そのような表情、つまり、そのような表情を引き起こす心理状態を検知することにより、異常事態についてより詳細に把握することができる。

[0050] 動作認識部107は、人物特定部102によって特定された人物、すなわち人物検索部105によって検索された人物の動作を認識する。具体的には、例えば、動作認識部107は、上述したアノテーションが付与された映像データに対して公知の動作認識処理を行って動作を認識する。例えば、動作認識部107は、画像に対して特許文献3で開示された技術などを用いて人

物の関節位置を追跡することにより、腕や手や脚の動き、姿勢などを特定する。特に、動作認識部１０７は、人物の動作が、所定の動作であるか否かについて分析する。ここで、この所定の動作は、予め登録された動作であって、例えば、異常事態に遭遇した人物が行う一連の動作であってもよいし、ジェスチャー（ポーズ）であってもよい。

[0051] 本実施の形態では、動作認識部１０７は、人物の動作を認識すると、異常行動記憶部１２３に記憶された一連の動作と、認識された動作とを照合することにより、当該人物が異常事態に遭遇した人物が行う動作を行ったか否かを判定する。すなわち、動作認識部１０７は、認識された人物の動作が予め定義された一連の動作と類似するか否かを分析する。動作認識部１０７は、両者の類似度が所定の閾値以上である場合、予め定義された一連の動作が実際に行われたと判定してもよい。

[0052] 異常行動記憶部１２３は、異常事態に遭遇した人物が行う一連の動作を表す情報を記憶するデータベースである。異常行動記憶部１２３に登録された一連の動作は、一つであってもよいし複数であってもよい。例えば、監視対象エリア９０が店舗である場合においては、店員が強盗に襲われた際には、店員がレジからお金を取り出して強盗に渡すことが考えられる。したがって、異常事態に遭遇した人物が行う一連の動作として、判定対象の人物（異常な音声を発した店員）の腕が、レジの方向に移動して、手で何かを取り出して、目の前の相手に差し出す動作を表す情報を異常行動記憶部１２３に記憶してもよい。

[0053] また、本実施の形態では、動作認識部１０７は、人物の動作を認識すると、ジェスチャー記憶部１２４に記憶されたジェスチャーと、認識された動作とを照合することにより、当該人物が異常事態に遭遇した人物が行う動作を行ったか否かを判定する。すなわち、動作認識部１０７は、認識された人物の動作（ジェスチャー）が予め定義されたジェスチャーと類似するか否かを分析する。動作認識部１０７は、両者の類似度が所定の閾値以上である場合、予め定義されたジェスチャーが実際に行われたと判定してもよい。

- [0054] ジェスチャー記憶部 124 は、異常事態に遭遇した際に行うよう従業員などに予め教育されたジェスチャーを表す情報を記憶するデータベースである。ジェスチャー記憶部 124 に登録されたジェスチャーは、一つであってもよいし複数であってもよい。例えば、監視対象エリア 90 が店舗である場合においては、店員などの従業員に対し、「強盗に襲われたら大声を出した上で、左手を上には伸ばすジェスチャーを行いながら、強盗の要求に従う」などの、取り決めや訓練を予め行っておく。この場合、左手を上には伸ばすジェスチャーを表す情報を異常行動記憶部 123 に予め記憶しておく。なお、登録するジェスチャーとしては、平素の従業員の行動では滅多になく、かつ、異常事態に遭遇した場合に不自然ではない動きであり、かつ、映像の分析により検知しやすいジェスチャーあることが望ましい。このように、ジェスチャーを事前に登録しておくことで、対応すべき異常事態を映像分析によって確実に検知することができる。
- [0055] 表情スコア算出部 108 は、表情認識部 106 が認識した表情についてのスコア値を算出する。表情スコア算出部 108 は、怒りや怯えなどの平常では起こりえない異常表情の異常度合いを数値化するスコアを算出する。例えば、表情スコア算出部 108 は、認識された表情が大きな怒り又は大きな怯えを表す表情であるほど大きな値を出力する。なお、表情の認識結果として笑顔度や怒り度などのように喜怒哀楽に関するスコア値が得られる場合、表情スコア算出部 108 は、認識処理で得られた怒り度などの所定の表情のスコア値を用いて、スコア値の出力を行ってもよい。
- [0056] 動作スコア算出部 109 は、動作認識部 107 が認識した動作についてのスコア値を算出する。本実施の形態では、動作スコア算出部 109 は、異常行動記憶部 123 に記憶された動作に関する異常度合いを数値化するスコアを算出する。例えば、動作スコア算出部 109 は、認識された一連の動作と異常行動記憶部 123 に記憶された動作の類似度が高いほど大きな値を出力する。動作スコア算出部 109 は、認識された動作が、予め定義された動作のいずれに該当するかに応じて、異なるスコア値を算出してもよい。なお、

本実施の形態では、動作スコア算出部 109 は、異常行動記憶部 123 に記憶された動作に関する異常度合いを数値化するスコアを算出するが、ジェスチャーについても同様にスコア値を算出してもよい。

[0057] 二次判定部 110 は、発生した異常事態に対して対応が必要であるか否かを判定する。具体的には、二次判定部 110 は、表情スコア算出部 108 及び動作スコア算出部 109 が算出したスコア値と、ジェスチャー記憶部 124 に定義されたジェスチャーが行われたか否かの判定結果とを用いて、対応が必要であるか否かを判定する。二次判定部 110 は、これらを入力として用いて、予め定められた判定ロジックに従って、対応が必要であるか否かを判定する。なお、二次判定部 110 は、これらの入力のうち一部だけを用いて判定を行ってもよい。例えば、二次判定部 110 は、表情スコア算出部 108 が算出されたスコア値が第 1 の閾値を超えた場合に、異常事態に対して対応が必要であると判定してもよい。また、二次判定部 110 は、動作スコア算出部 109 が算出されたスコア値が第二の閾値を超えた場合に、異常事態に対して対応が必要であると判定してもよい。また、二次判定部 110 は、算出された 2 つのスコア値の合計が第 3 の閾値を超えた場合に異常事態に対して対応が必要であると判定してもよい。また、二次判定部 110 は、予め定義されたジェスチャーが行われた場合に、異常事態に対して対応が必要であると判定してもよい。また、二次判定部 110 は、予め定義されたジェスチャーが行われたか否かによって、上述した閾値の値を変更してもよい。すなわち、予め定義されたジェスチャーが行われた場合、予め定義されたジェスチャーが行われていない場合よりも低い閾値が用いられてもよい。なお、上述した判定ロジックは例に過ぎず、二次判定部 110 は、任意の判定ロジックを用いることができる。このように、本実施の形態では、二次判定部 110 は、映像分析の結果に基づいて、監視対象エリア 90 における異常事態を評価する。

[0058] 信号出力部 111 は、発生した異常事態に対して対応が必要であると二次判定部 110 によって判定された場合に、異常事態に対して対応するための

所定の信号を出力する。すなわち、信号出力部 111 は、異常事態の評価が所定の基準を満たす場合に、所定の信号を出力する。この所定の信号は、他のプログラム（他の装置）もしくは人間に対して、所定の指示をするための信号であってもよい。例えば、所定の信号は、警備員室などで警報ランプと警報音を発報させるための信号であってもよいし、警備員等へ異常事態に対する対応を指示するメッセージであってもよい。また、所定の信号は、犯罪的な行為を抑制すべく、異常事態の発生位置の付近の警告灯を点滅させるための信号であってもよいし、異常事態の発生位置の近辺にいる人々へ避難を促す警報の出力のための信号であってもよい。

[0059] 図4に示した機能及び図5に示した機能は、例えば図6に示すようなコンピュータ50により実現されてもよい。図6は、コンピュータ50のハードウェア構成の一例を示す模式図である。図6に示すように、コンピュータ50は、ネットワークインタフェース51、メモリ52、及びプロセッサ53を含む。

[0060] ネットワークインタフェース51は、他の任意の装置と通信するために使用される。ネットワークインタフェース51は、例えば、ネットワークインタフェースカード（NIC）を含んでもよい。

[0061] メモリ52は、例えば、揮発性メモリ及び不揮発性メモリの組み合わせによって構成される。メモリ52は、プロセッサ53により実行される、1以上の命令を含むプログラム、及び各種処理に用いるデータなどを格納するために使用される。

[0062] プロセッサ53は、メモリ52からプログラムを読み出して実行することで、図4又は図5に示した各構成要素の処理を行う。プロセッサ53は、例えば、マイクロプロセッサ、MPU(Micro Processor Unit)、又はCPU(Central Processing Unit)などであってもよい。プロセッサ53は、複数のプロセッサを含んでもよい。

[0063] プログラムは、コンピュータに読み込まれた場合に、実施形態で説明された1又はそれ以上の機能をコンピュータに行わせるための命令群（又はソフ

トウェアコード)を含む。プログラムは、非一時的なコンピュータ可読媒体又は実体のある記憶媒体に格納されてもよい。限定ではなく例として、コンピュータ可読媒体又は実体のある記憶媒体は、random-access memory (RAM)、read-only memory (ROM)、フラッシュメモリ、solid-state drive (SSD)又はその他のメモリ技術、CD-ROM、digital versatile disc (DVD)、Blu-ray (登録商標)ディスク又はその他の光ディスクストレージ、磁気カセット、磁気テープ、磁気ディスクストレージ又はその他の磁気ストレージデバイスを含む。プログラムは、一時的なコンピュータ可読媒体又は通信媒体上で送信されてもよい。限定ではなく例として、一時的なコンピュータ可読媒体又は通信媒体は、電氣的、光学的、音響的、またはその他の形式の伝搬信号を含む。

[0064] 次に、監視システム10の動作の流れについて説明する。図7は、監視システム10の動作の流れの一例を示すフローチャートである。また、図8は、図7に示したフローチャートにおけるステップS107における処理の流れの一例を示すフローチャートである。

[0065] 以下、図7及び図8に沿って、監視システム10の動作の流れの一例について説明する。本実施の形態では、ステップS101及びステップS102が、音響センサ300の処理として実行され、ステップS103以降の処理が分析サーバー100の処理として実行される。

[0066] ステップS101において、異常検知部301が、音響センサ300が検知した音に基づいて、監視対象エリア90内での異常事態の発生を検知する。

[0067] 次に、ステップS102において、一次判定部302は、発生した異常事態に対して対応が不要であるか否かを判定する。発生した異常事態に対して対応が不要であると判定された場合(ステップS102でYes)、処理はステップS101に戻り、そうでない場合(ステップS102でNo)、処理はステップS103へ移行する。

[0068] ステップS103において、音声取得部101は、音響センサ300から

所定の音声を取得し、人物特定部１０２は、音声取得部１０１が取得した所定の音声から得られる特徴に基づいて、この所定の音声を発声した人物を特定する。

[0069] 次に、ステップＳ１０４において、音源位置推定部１０３は、音響センサ３００の出力に基づいて、所定の音声の音源位置（異常事態の発生位置）を推定する。

[0070] 次に、ステップＳ１０５において、映像の分析のために、映像取得部１０４が、監視対象エリア９０に設けられた全ての監視カメラ２００のうち、異常事態の発生位置を撮影している監視カメラ２００から、映像データを取得する。このため、複数の監視カメラ２００のうち、異常事態の発生位置を含むエリア（音源位置を含むエリア）を撮影する監視カメラ２００の映像データに対してだけ、分析処理が行われる。

[0071] また、映像を構成する画像内の部分画像であって、音源位置を含む部分画像に対してだけ、分析処理が行われてもよい。すなわち、音源位置を含むエリアを撮影する監視カメラ２００の撮影領域全体の画像ではなく、一部分の領域に相当する部分画像に対してだけ、分析処理が行われてもよい。

[0072] また、本実施の形態では、映像分析の処理は、平常時には実行されず、異常事態が発生した場合にしか実行されない。つまり、監視カメラ２００の映像を用いた分析処理は、異常事態の発生が検知された場合（具体的には、所定の音声が発検知された場合）に実行され、異常事態の発生が発検知される前（所定の音声が発検知される前）には実行されない。

[0073] 次に、ステップＳ１０６において、人物検索部１０５が、人物の外観の特徴に基づいて、ステップＳ１０５で取得された映像から、人物特定部１０２により特定された人物を検索する。

[0074] 次に、ステップＳ１０７において、検索された人物についての映像分析が行われる。ステップＳ１０７の処理について、図８を参照して、具体的に説明する。映像分析では、まず、ステップＳ２０１及びステップＳ２０３の処理が行われる。なお、ステップＳ２０１及びその後続の処理と、ステップＳ

203及びその後続の処理は、例えば並行に実行されるが、シーケンシャルに実行されてもよい。

[0075] ステップS201では、表情認識部106が、ステップS106で検索された人物の表情を認識する。ステップS201の後、ステップS202において、表情スコア算出部108が、ステップS201での認識結果に基づいて異常表情についてのスコア値を算出する。

[0076] これに対し、ステップS203では、動作認識部107が、ステップS106で検索された人物の動作を認識する。ステップS203の後、ステップS204において、動作認識部107は、ジェスチャー記憶部124に記憶されたジェスチャーが検知されたか否かを確認する。また、ステップS203の後、ステップS205において、動作スコア算出部109は、ステップS203の認識結果に基づいて、異常行動記憶部123に記憶された動作に関するスコア値を算出する。

[0077] ステップS202、S204、及びS205の処理が終了すると、処理は図7に示したステップS108へ移行する。

[0078] ステップS108において、二次判定部110が、ステップS107の処理結果に基づいて、発生した異常事態に対して対応が不要であるか否かを判定する。発生した異常事態に対して対応が不要であると判定された場合（ステップS108でYes）、処理はステップS101に戻り、そうでない場合（ステップS108でNo）、処理はステップS109へ移行する。

[0079] ステップS109において、信号出力部111は、異常事態に対して対応するための所定の信号を出力する。これにより、異常事態に対して対応することが可能になる。ステップS109の後、処理はステップS101に戻る。

[0080] 以上、実施の形態について説明した。監視システム10によれば、上述したとおり、音声及び映像を用いた処理が行われ、これにより、異常事態の発生を検知し、異常事態を適切に把握することができる。

[0081] 特に、監視システム10では、まずは人が発する異常な音声から異常事態

の発生が検知され、当該音声の特徴から異常な音声を発した人物が特定される。その上で、監視システム１０は、映像に基づいて、異常な音声を発した人物の表情や行動を分析することにより、異常事態の発生について詳細な確認処理を実施する。このように、本実施の形態では、映像に対する分析は、異常な音声の検知とともに行われる。その理由は、犯罪や事故の種類は千差万別なため、何らかの前提条件が加わらない限り、不測の異常事態に対する映像特徴を事前に定義するのは困難であるからである。すなわち、「異常な音声を発した人物」という前提条件が付加されると、映像に映った当該の人物の表情や行動から、異常事態の発生を確認することが容易である。つまり、そのような前提条件があることにより、例えば、お客から代金を受け取ってレジから釣銭を渡す行動と、強盗から脅されてレジから現金を渡す行動とを容易に区別することができる。

[0082] また、音と映像を用いた処理が行われる利点として次のような利点もある。音を分析することによる異常事態の発生の検知は、不測の異常事態に対しても有効であるが、音による分析だけでは、検知した異常事態に対して対応が必要であるか否かなどの評価をすることは難しい。音による異常検知は、人が目を瞑って耳を澄ましている場合と同様であり、悲鳴や叫び声の検知により、異常事態が発生した可能性が高いということまでは認識できても、それ以上の詳細な状況を把握することはできない。従って、例えば直ちに警備員等を急行させるべきか、あるいは翌日まで待ってから確認すれば良い程度の軽微な異常なのかなど、音からは異常事態の詳細を把握することが困難である。これに対し、異常な音声を発した人物の表情や行動の映像分析を加えることで、異常事態について詳細に評価することが可能となる。このように、本実施の形態では、まずは人が発する異常な音声から異常事態の発生を検知し、異常な音声を発した人物を特定した上で、当該人物の映像から表情や行動を分析するという、音と映像によるマルチモーダルな分析を実現している。

[0083] また、上述したとおり、分析サーバー１００における映像の分析処理は、

異常な音声の音源位置の近傍の映像に対してだけ行われてもよい。すなわち、複数の監視カメラ２００の映像のうち、音源位置だと推定された位置を撮影している監視カメラ２００の映像に対してだけ分析が行われてもよい。また、１台の監視カメラ２００の映像の中から切り出された、音源位置だと推定された位置を含む部分画像に対してだけ分析が行われてもよい。映像をリアルタイムに分析するには、多大なコンピュータ資源を必要とする。しかしながら、本実施の形態では、音源位置の近傍の映像のみを分析対象とすることにより、コンピュータ資源の利用を抑制することができる。また、本実施の形態によれば、上述した通り、映像分析の処理は、平常時には実行されず、音による異常事態の検知が行われた場合に限って実行される。このため、本実施の形態によれば、コンピュータ資源の利用をさらに抑制することができる。

[0084] ＜実施の形態の変形例＞

上述した実施の形態では、音響センサ３００が配置され、音響センサ３００が異常検知部３０１及び一次判定部３０２を備えたが、このような構成に代えて、次のような構成により監視システムが構成されてもよい。すなわち、音響センサ３００の代わりにマイクを監視対象エリア９０に配置し、マイクで集音した音声信号を分析サーバー１００に伝送して、分析サーバー１００が音響分析や音声認識を行ってもよい。つまり、音響センサ３００の構成要素のうち、少なくともマイクだけが、監視対象エリア９０に配置されていればよく、その他の構成要素については、監視対象エリア９０に配置されていなくてもよい。このように、上述した異常検知部３０１及び一次判定部３０２の処理が分析サーバー１００により実現されてもよい。

[0085] なお、上述した実施の形態に示される監視方法を、監視プログラムとして実装して販売してもよい。この場合、ユーザは任意のハードウェア上にそれをインストールし、利用することができるため、利便性が向上する。また、上述した実施の形態に示される監視方法を、監視装置として実装してもよい。この場合、ユーザは自分でハードウェアを準備してプログラムをインストール

ールする手間がかからずに上述した監視方法を利用することができるため、利便性が向上する。また、上述した実施の形態に示される監視方法を、複数の装置により構成したシステムとして実装してもよい。この場合、ユーザは自分で複数の装置を組み合わせて調整する手間がかからず、上述した監視方法を利用することができるため、利便性が向上する。

[0086] 以上、実施の形態を参照して本願発明を説明したが、本願発明は上記によって限定されるものではない。本願発明の構成や詳細には、発明のScope内で当業者が理解し得る様々な変更をすることができる。

[0087] 上記の実施形態の一部又は全部は、以下の付記のようにも記載され得るが、以下には限られない。

(付記1)

監視対象エリアにおける異常事態の発生により人が発声した所定の音声を取得する音声取得手段と、

前記所定の音声から得られる特徴に基づいて、前記所定の音声を発声した人物を特定する人物特定手段と、

前記監視対象エリアを撮影するカメラの映像から、特定された前記人物を検索し、当該人物の表情又は動作を分析する分析手段と、

前記分析の結果に基づいて、前記監視対象エリアにおける前記異常事態を評価する異常事態評価手段と

を有する監視装置。

(付記2)

前記分析手段は、前記人物の表情が所定の表情であるか否かを分析する

付記1に記載の監視装置。

(付記3)

前記分析手段は、前記人物の動作が予め定義された一連の動作と類似するか否かを分析する

付記1又は2に記載の監視装置。

(付記4)

前記分析手段は、前記人物の動作が予め定義されたジェスチャーと類似するか否かを分析する

付記 1 乃至 3 のいずれか一項に記載の監視装置。

(付記 5)

前記分析手段による分析処理が、前記所定の音声を検知された場合に実行され、前記所定の音声を検知される前には実行されない

付記 1 乃至 4 のいずれか一項に記載の監視装置。

(付記 6)

前記所定の音声の音源位置を推定する音源位置推定手段をさらに有し、
前記分析手段は、複数の前記カメラのうち、前記音源位置を含むエリアを撮影するカメラの映像データに対してだけ、分析処理を行う

付記 1 乃至 5 のいずれか一項に記載の監視装置。

(付記 7)

前記分析手段は、前記映像を構成する画像内の前記音源位置を含む部分画像に対してだけ、分析処理を行う

付記 6 に記載の監視装置。

(付記 8)

前記異常事態の評価が所定の基準を満たす場合に、所定の信号を出力する信号出力手段をさらに有する

付記 1 乃至 7 のいずれか一項に記載の監視装置。

(付記 9)

監視対象エリアを撮影するカメラと、

監視対象エリアにおいて発生する音を検知するセンサと、

監視装置と

を備え、

前記監視装置は、

前記監視対象エリアにおける異常事態の発生により人が発声した所定の音声を前記センサから取得する音声取得手段と、

前記所定の音声から得られる特徴に基づいて、前記所定の音声を発声した人物を特定する人物特定手段と、

前記カメラの映像から、特定された前記人物を検索し、当該人物の表情又は動作を分析する分析手段と、

前記分析の結果に基づいて、前記監視対象エリアにおける前記異常事態を評価する異常事態評価手段と

を有する、

監視システム。

(付記 10)

前記分析手段は、前記人物の表情が所定の表情であるか否かを分析する

付記 9 に記載の監視システム。

(付記 11)

前記分析手段は、前記人物の動作が予め定義された一連の動作と類似するか否かを分析する

付記 9 又は 10 に記載の監視システム。

(付記 12)

前記分析手段は、前記人物の動作が予め定義されたジェスチャーと類似するか否かを分析する

付記 9 乃至 11 のいずれか一項に記載の監視システム。

(付記 13)

監視対象エリアにおける異常事態の発生により人が発声した所定の音声を取得し、

前記所定の音声から得られる特徴に基づいて、前記所定の音声を発声した人物を特定し、

前記監視対象エリアを撮影するカメラの映像から、特定された前記人物を検索し、当該人物の表情又は動作を分析し、

前記分析の結果に基づいて、前記監視対象エリアにおける前記異常事態を評価する

監視方法。

(付記 14)

監視対象エリアにおける異常事態の発生により人が発声した所定の音声を取得する音声取得ステップと、

前記所定の音声から得られる特徴に基づいて、前記所定の音声を発声した人物を特定する人物特定ステップと、

前記監視対象エリアを撮影するカメラの映像から、特定された前記人物を検索し、当該人物の表情又は動作を分析する分析ステップと、

前記分析の結果に基づいて、前記監視対象エリアにおける前記異常事態を評価する異常事態評価ステップと

をコンピュータに実行させるプログラムが格納された非一時的なコンピュータ可読媒体。

符号の説明

- [0088] 1 監視装置
- 2 音声取得部
- 3 人物特定部
- 4 分析部
- 5 異常事態評価部
- 10 監視システム
- 50 コンピュータ
- 51 ネットワークインタフェース
- 52 メモリ
- 53 プロセッサ
- 90 監視対象エリア
- 100 分析サーバー
- 101 音声取得部
- 102 人物特定部
- 103 音源位置推定部

1 0 4	映像取得部
1 0 5	人物検索部
1 0 6	表情認識部
1 0 7	動作認識部
1 0 8	表情スコア算出部
1 0 9	動作スコア算出部
1 1 0	二次判定部
1 1 1	信号出力部
1 2 1	音声特徴記憶部
1 2 2	外観特徴記憶部
1 2 3	異常行動記憶部
1 2 4	ジェスチャー記憶部
2 0 0	監視カメラ
3 0 0	音響センサ
3 0 1	異常検知部
3 0 2	一次判定部
5 0 0	ネットワーク

請求の範囲

- [請求項1] 監視対象エリアにおける異常事態の発生により人が発声した所定の音声を取得する音声取得手段と、
前記所定の音声から得られる特徴に基づいて、前記所定の音声を発声した人物を特定する人物特定手段と、
前記監視対象エリアを撮影するカメラの映像から、特定された前記人物を検索し、当該人物の表情又は動作を分析する分析手段と、
前記分析の結果に基づいて、前記監視対象エリアにおける前記異常事態を評価する異常事態評価手段と
を有する監視装置。
- [請求項2] 前記分析手段は、前記人物の表情が所定の表情であるか否かを分析する
請求項1に記載の監視装置。
- [請求項3] 前記分析手段は、前記人物の動作が予め定義された一連の動作と類似するか否かを分析する
請求項1又は2に記載の監視装置。
- [請求項4] 前記分析手段は、前記人物の動作が予め定義されたジェスチャーと類似するか否かを分析する
請求項1乃至3のいずれか一項に記載の監視装置。
- [請求項5] 前記分析手段による分析処理が、前記所定の音声が発知された場合に実行され、前記所定の音声が発知される前には実行されない
請求項1乃至4のいずれか一項に記載の監視装置。
- [請求項6] 前記所定の音声の音源位置を推定する音源位置推定手段をさらに有し、
前記分析手段は、複数の前記カメラのうち、前記音源位置を含むエリアを撮影するカメラの映像データに対してだけ、分析処理を行う
請求項1乃至5のいずれか一項に記載の監視装置。
- [請求項7] 前記分析手段は、前記映像を構成する画像内の前記音源位置を含む

部分画像に対してだけ、分析処理を行う

請求項 6 に記載の監視装置。

[請求項8] 前記異常事態の評価が所定の基準を満たす場合に、所定の信号を出力する信号出力手段をさらに有する

請求項 1 乃至 7 のいずれか一項に記載の監視装置。

[請求項9] 監視対象エリアを撮影するカメラと、
監視対象エリアにおいて発生する音を検知するセンサと、
監視装置と
を備え、
前記監視装置は、
前記監視対象エリアにおける異常事態の発生により人が発声した所定の音声を前記センサから取得する音声取得手段と、
前記所定の音声から得られる特徴に基づいて、前記所定の音声を発声した人物を特定する人物特定手段と、
前記カメラの映像から、特定された前記人物を検索し、当該人物の表情又は動作を分析する分析手段と、
前記分析の結果に基づいて、前記監視対象エリアにおける前記異常事態を評価する異常事態評価手段と
を有する、
監視システム。

[請求項10] 前記分析手段は、前記人物の表情が所定の表情であるか否かを分析する

請求項 9 に記載の監視システム。

[請求項11] 前記分析手段は、前記人物の動作が予め定義された一連の動作と類似するか否かを分析する

請求項 9 又は 10 に記載の監視システム。

[請求項12] 前記分析手段は、前記人物の動作が予め定義されたジェスチャーと類似するか否かを分析する

請求項 9 乃至 11 のいずれか一項に記載の監視システム。

- [請求項13] 監視対象エリアにおける異常事態の発生により人が発声した所定の音声を取得し、
- 前記所定の音声から得られる特徴に基づいて、前記所定の音声を発声した人物を特定し、
- 前記監視対象エリアを撮影するカメラの映像から、特定された前記人物を検索し、当該人物の表情又は動作を分析し、
- 前記分析の結果に基づいて、前記監視対象エリアにおける前記異常事態を評価する
- 監視方法。

- [請求項14] 監視対象エリアにおける異常事態の発生により人が発声した所定の音声を取得する音声取得ステップと、
- 前記所定の音声から得られる特徴に基づいて、前記所定の音声を発声した人物を特定する人物特定ステップと、
- 前記監視対象エリアを撮影するカメラの映像から、特定された前記人物を検索し、当該人物の表情又は動作を分析する分析ステップと、
- 前記分析の結果に基づいて、前記監視対象エリアにおける前記異常事態を評価する異常事態評価ステップと
- をコンピュータに実行させるプログラムが格納された非一時的なコンピュータ可読媒体。