筑波大学大学院博士課程

システム情報工学研究科修士論文

# ライフログ画像から得られる人物、物体情報を用いた検索・閲覧システムの構築

平馬　克彦

修士（工学）

（コンピュータサイエンス専攻）

指導教員　田中　二郎

2016 年　3 月

# 概要

ライフログ(Life Log)は、人間の生活・行い・体験(Life)を、長期間に渡り記録(Log)を行うこと、または記録自体のことを指す。

本論文では、まずライフログの定義と現行のライフログデバイスの調査を行った。ライフログは記録の継続により、ユーザが必要とするログデータを発見することが困難になっていくこと、現行のライフログ画像検索・閲覧システムでは、ユーザが自身の過去の分析などが十分に行えない問題を示した。この問題点を解決するため、ライフログ画像の情報を用いた検索・閲覧システムの構築を本論文の目的とした。そこで、ライフログ閲覧の活用である記憶想起に着目し、記憶想起における人物・物体情報の重要性を述べた。それらより、人物・物体情報をライフログ画像から取得し、検索・閲覧に用いる想定システムの提案を行った。想定システムを踏まえ、顔検出による人物情報の取得、物体検出による物体情報の取得、画像を入力に用いたライフログ画像検索によるプロトタイプシステムを提案した。プロトタイプシステムの部分実装として、顔検出と物体検出、画像認識マッチング処理実装を行った。

# 目次

第 1 章　ライフログとは ……1
1.1　ライフログの定義 ……1
1.2　ライフログ収集デバイス ……2
1.2.1　Vicon Revue ……2
1.2.2　Narrative clip ……4
1.2.3　その他のデバイス ……5
第 2 章　現行のライフログ閲覧システム ……6
2.1　Vicon Revue Desktop ……6
2.2　Narrative Clip 閲覧システム ……8
第 3 章　ライフログにおける検索・閲覧の重要性 ……10
第 4 章　研究目的 ……11
第 5 章　想定システム ……12
第 6 章　プロトタイプシステムの部分実装 ……15
6.1　概要 ……15
6.2　顔検出 ……16
6.2.1　概要 ……16
6.2.2　Haar-Like 特徴量 ……17
6.2.3　カスケード分類器 ……17
6.3　物体検出 ……19
6.3.1　概要 ……19
6.3.2　Selective Search ……19
6.3.3　Labellio ……20
6.4　画像特徴量マッチング ……21
6.4.1　概要 ……21
6.4.2　SIFT (Scale-Invariant Feature Transform) ……21
6.4.3　対応点探索による画像マッチング ……26
第 7 章　関連研究 ……27
7.1　ライフログの検索・閲覧に関する研究 ……27
7.2　記憶想起に関する研究 ……28
第 8 章　まとめ ……29
謝辞 ……30
参考文献 ……31

# 図目次

図 1-1 Vicon Revue 撮影画像 ..... 3
図 1-2 Narrative clip 撮影画像 ..... 4
図 2-1 Vicon Revue アルバム一覧 ..... 7
図 2-2 Vicon Revue アルバム閲覧 ..... 7
図 2-3 Narrative Clip アルバム表示 ..... 9
図 2-4 Narrative Clip リスト表示 ..... 9
図 5-1 想定システム概要 ..... 13
図 5-2 検索結果閲覧イメージ図 ..... 14
図 5-3 画像拡大認識イメージ図 ..... 14
図 6-1 プロトタイプシステム概要図 ..... 15
図 6-2 顔検出前ライフログ画像 ..... 16
図 6-3 顔検出後ライフログ画像 ..... 16
図 6-4 Haar-Like 矩形パターン例 ..... 17
図 6-5 カスケード分類器アルゴリズム ..... 18
図 6-6 SelectiveSearch 処理後画像 ..... 19
図 6-7 Labellio の画像認識モデル適用結果 ..... 20

# 表目次

表 1-1 記録法別ライフログの利点と欠点 ..... 1

# 第1章　ライフログとは

モバイル機器、ウェアラブルデバイスの普及[1]により、日常の情報を様々な形で取得する事が可能になってきた。機器から得られた情報を分析、利用するためのアプリケーションやサービスが普及しつつある。同時に、日常のログデータは記録を継続することで、分析に有用になるが、ログデータの量は膨大になっていく。

## 1.1　ライフログの定義

ライフログ(Life Log)とは、人間の生活・行い・体験(Life)を、長期間に渡り記録(Log)を行うこと、または記録自体のことを指す[2]。全ての操作・動作を後から追跡し、それによって過去の分析や追体験を可能にするものである。広義的には、日記やネットブログの類も、ライフログと呼ばれることがある。

ライフログには、ユーザが自分で操作して記録する手動記録と，外部デバイスにより自動的に記録する自動記録がある。 前者の手動記録は、詳細で自由度の高い記録が可能であり、ブログやメモなどのように、記録にユーザの主観的意見を含めることができるが、ユーザ操作を伴うため、記録に対するユーザの負担が大きい。さらに、ユーザの記録忘れなどから、ライフログに欠落が出る場合がある。後者の自動記録は、カメラや加速度センサ、GPS などのセンサ類が搭載されているウェアラブルデバイスを装着して、画像・動画・音声・位置情報といったデータを常時記録するというものである。 ユーザの記録負担は小さい。取得されるデータはセンサによって限定されており、客観的なデータを取得できる。表 1 は、手動記録ライフログと自動記録ライフログの利点、欠点を表にしたものである。

| | 手動記録ライフログ | 自動記録ライフログ |
|---|---|---|
| 利点 | 自由度の高い記録が可能<br>主観的な内容を記録可能 | ユーザに負担がかかりにくい<br>異なるデータを同時に常時記録可能 |
| 欠点 | ユーザに負担がかかる<br>記録に欠落がでる | 主観的内容を取得しづらい<br>数値などを分析する必要がある |

**表 1-1 記録法別ライフログの利点と欠点**

近年では、ウェアラブルデバイスの世界市場成長により、デバイスの普及が著しく伸びている。それにより、ウェアラブルデバイスを用いた自動的なデータ収集は容易になりつつある。本研究では、記録にユーザの負担がかかりにくく、複数のデータを常時記録が可能である自動記録ライフログを扱う。

ライフログの日常体験記録という発想の原型は、1945 年に米国の科学技術政策の転換を促した技術者 Vannervar Bush が提唱したシステム「Memex」である[3][4]。Memex は、個人が所有する全ての本、記録、通信内容などを圧縮して格納できるシステムであり、高速かつ柔軟に参照できるように機械化されているものである。個人が使用する電気機械式装置として描かれており、大きな自己完結型の図書館を構成している。Memex では、個人の蔵書や記録などをすべて保存した後に、人間の記憶システムと同じように、データを互いに関連付けることで「個人の記憶を拡張する個人的な補助記憶」を提供している。

その後の代表的な研究として、Gemmell らが 2002 年から行った「MyLifeBits Project」[5][6]がある。MyLifeBits は、コンピュータを使用する際に行われる全ての操作・動作を、後から追跡し、それによって過去の分析や追体験を可能にするものである。MyLifeBits では、メール、 閲覧した Web サイト、紙書類のスキャンデータ、 画像、音声データ、本、音楽 CD を対象としている。それらを記録するのに必要な記録媒体の一人あたりの容量は、5 年間で約 80GB になると試算されている。

代表的なライフログ収集用デバイスとして、William らが開発した SenseCam[7]がある。SenseCam には魚眼レンズや加速度センサ、赤外線モーションセンサなどのセンサ類が搭載されている。首から下げる形で SenseCam を装着してもらい、一人称視点の写真を撮影する事ができる。SenseCam は、疾患または事故などによる脳外傷より、記憶障害を患った患者への支援を目的に作成された。撮影された画像を検討することによって、アルツハイマー病、記憶喪失、およびその他の記憶障害を患っている患者が、失われた記憶を取得することが判明している。

## 1.2 ライフログ収集デバイス

### 1.2.1 Vicon Revue

Vicon Revue は、Micorosoft Research が研究している SenseCam 技術に基づいたライフログ取得用のウェラブルカメラである。首からぶら下げておくことで、30 秒ごとに自動的に写真を撮影する。さらに内蔵されている多様なセンサ(照度センサ、赤外線モーションセンサ、加速度センサ)によって、外界の変化が検出されたときにも撮影される。基本的に、30 秒間に 1 回撮影を行うのみだが、センサが動きを探知した場合にも 1 回撮影される。歩く事や、日常的に使用した場合、1 分間に 5 枚程撮影するようになる。他の機能として、シャッターボタンとプライバシーボタンがある。シャッターボタンを押す事で、自主的に撮影する事ができ、自主的に撮影した事も保存される。プライバシーボタンは、一定時間撮影を停止する事ができる。この非撮影状態は、プライバシーボタンを再度押すことで延長する事ができる。およそ 12 日間の写真を蓄積でき、添付アプリケーションを用いることで日々アップロードできる。撮りためた写真は、添付アルバムアプリケーションで閲覧ができる。Vicon Revue は、記憶障害のある患者に対する支援のために使用されている。図 1-1 は、実際に Vicon Revue で撮影された画像である。

図 1-1 Vicon Revue 撮影画像

### 1.2.2 Narrative clip

Narrative clip[8]は、Narrative が製作、販売をしているライフログ収集用ウェアラブルカメラである。Sensecam に比べて、非常に小型になっており、後ろにクリップが備え付けられている。ユーザは首もとのどこかに装着する事で、画像を撮影する事ができる。30 秒毎の自動的な撮影と、Narrative clip 本体を二回タップする事で自主的な撮影の 2 種類の撮影方法がある。搭載されているセンサは、加速度センサと GPS のみであり、外界の変化によって撮影する事はできない。ON/OFF するためのボタンは存在せず、ポケットなどの暗い場所に入れるか、机の上でカメラを下向きにして置くことで非撮影状態にすることができる。約 8000 枚の写真が撮影可能である。撮影された画像は、コンピュータと USB 接続する事で、専用のクラウドサーバーへ送信される。このとき、ローカル上に保存する事も可能である。送信された画像は、ウェブブラウザまたはスマートフォンの専用アプリケーションから閲覧する事ができる。図 1-2 は、NarrativeClip で撮影された画像である。

図 1-2 Narrative clip 撮影画像

### 1.2.3 その他のデバイス

この他に、加速度センサやカラーセンサなどのセンサ情報を重視したライフログ収集デバイス「Autographer」[9]がある。Autographer は、定期的に撮影を行う事はせず、センサ情報から画像を撮影するタイミングをシステムが判断し、撮影する。撮影するタイミングを調整する事で、無用な画像を撮影しないようにしている。

その他にも、手首に装着し、加速度センサなどから使用者の活動量を取得する Fitbit や、眼電位センサから、使用者の視線などを取得可能な眼鏡型デバイスの JINS Meme など、ライフログを画像として取得しないデバイスも存在する。

# 第2章　現行のライフログ閲覧システム

## 2.1　Vicon Revue Desktop

Vicon revue が提供する「Vocon Revue Desktop」では、Vicon revue をパソコンと USB 接続することで、写真を閲覧システムにアップロードすることができる。アップロードされた写真は、日時に関わらず Vicon revue が起動した毎に生成されるアルバムに保存される。そのため、日を跨いで記録を行った場合も一つのアルバムに入ってしまうため、日ごとの検索が難しい。アルバムには、情報として「撮影開始時刻」「撮影時間」「アップロード日時」「最終閲覧時刻」「閲覧回数」が保存されている。この他に、ユーザがアルバム毎に「タイトル」「タグ」を手動で記録する事ができる。

アルバムを選択すると、アルバム内にある画像一枚一枚を閲覧する事ができる。画像閲覧ウィンドウには、時間軸に応じたシークバーがある。このシークバーを操作する事で、ユーザは他の画像を閲覧する事ができる。シークバーには、撮影開始時刻と終了時刻が、シークバーの端に表示され、シークバー中央に現在閲覧中の画像の撮影時刻が表示される。アルバム中の画像は、スライドショーとして閲覧する事も可能である。スライドショーの再生スピードを変更する事ができ、ユーザに合わせて再生することができる。表示している画像に対して、ユーザは「ノート」と呼ばれる自由記述型のメモを記述する事ができる。この「ノート」には、ユーザが記述した内容以外に、「ノートを記述した画像の撮影時刻」が記録され、シークバーには、時間に応じた場所に「ノート」が書かれたことを表すアイコンが表示される。同時に、「ノート」の中には、「ノート」を記述した画像を瞬時に表示するためのジャンプ機能がある。Vicon revue が首から下げて撮影を行うため、いくつかの写真に大きなブレが生じてしまい、ライフログとして活用できない写真も見受けられた。

Vicon Revue Desktop のアルバム一覧画面には、検索欄がある。ユーザが記述した「タイトル」「タグ」についてのみ検索をすることができる。日時に関しては、「タイトル」「タグ」に記述しない限り、検索することができない。「ノート」は、各アルバム内を閲覧する際に一覧で確認する事ができるが、「ノート」の内容に対して、検索を行うことはできない。同時に、アルバム一覧画面にある検索は、アルバムの「タイトル」「タグ」に対するものであり、「ノート」の内容の検索は不可能である。図 2-1 は、Vicon Revue デスクトップのアルバム一覧を表示したものである。図 2-2 は、あるアルバムを選択し、アルバム内の画像を閲覧している際のものである。

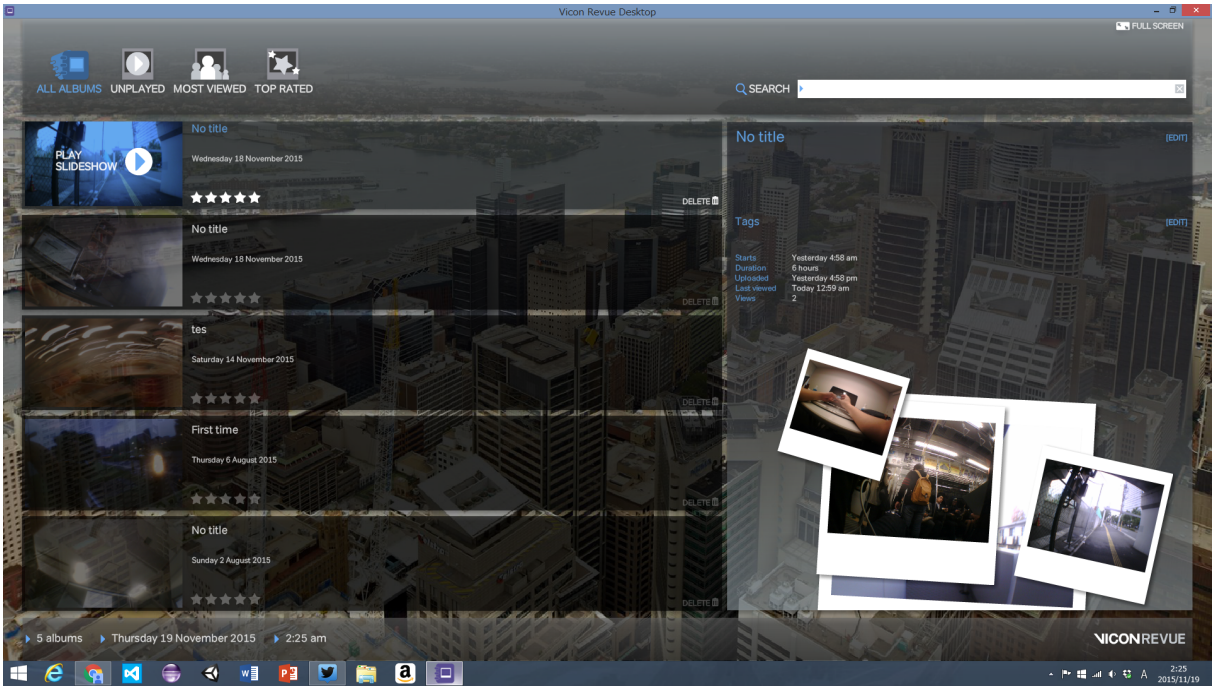



図 2-1 Vicon Revue アルバム一覧

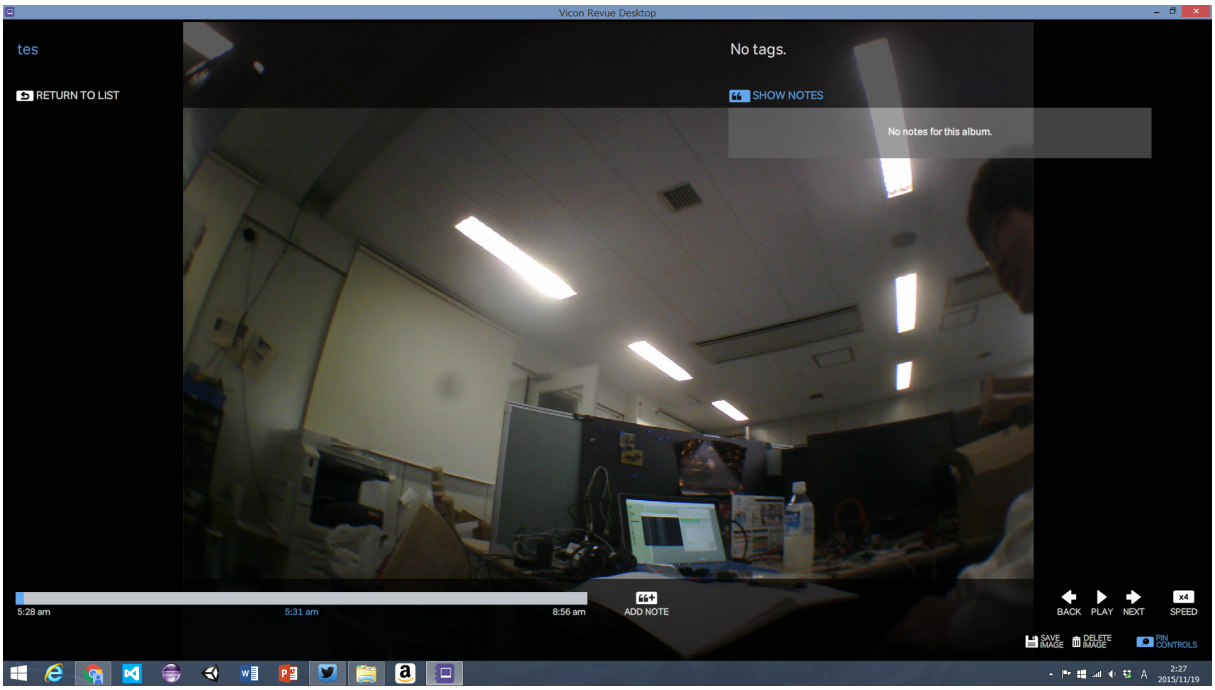



図 2-2 Vicon Revue アルバム閲覧

## 2.2 Narrative Clip 閲覧システム

Narrative Clip が提供する閲覧システムでは、Narrative Clip とパソコンを USB 接続する事で、写真を閲覧システムにアップロードすることができる。この際に、Narrative Clip が用意しているクラウド上だけでなく、USB 接続しているパソコンのローカル上に保存することができる。アップロードされた画像は、パソコンまたはスマートフォンで閲覧する事ができる。アップロードされた画像は、起動し、撮影される毎にアルバムとして保存される。しかし、前後のアルバム同士を統合する事や、逆に一つのアルバムを二つのアルバムに分割する事も可能である。アルバムには「撮影日」「アルバム内にある画像の撮影時刻」が表示されている。各アルバムには、140 字以内のキャプションを記録する事ができる。アルバムはアプリケーション内または Facebook などの SNS を通じて、他者と共有することができる

アルバムを選択すると、アルバム内にある画像のサムネイルリストが表示される。画像リストは、アルバム内の全画像表示と、システムが有用であると判断した画像表示の 2 種類のどちらかを選択することができる。リストの情報として「撮影日時」だけでなく、リスト内の画像枚数も表示される。リスト表示されている画像の枚数ある画像をフォーカスすると、その画像が手動で記録されているかどうかや、お気に入りとして登録されているか、確認する事ができる。同時に、画像を複数選択する事ができ、選択された画像を全てお気に入り、エクスポート、消去することができる。

リスト内の画像を選択する事で、その画像のみを表示する事ができる。選択した画像の撮影時間や撮影場所は閲覧する事ができない。画像に対して、任意の場所に「タグ」を追加する事ができる。「タグ」は、人物情報についてのみ付与することが可能である。人物情報は、同じく Narrative clip を使用している人物の情報を付加、利用する事ができる。

しかし、記述型のキャプションや、人物情報を付与できる「タグ」があるにも関わらず、Narrative clip の閲覧システムには検索システムが存在していない。ユーザはアルバム毎の情報を、検索システムなしで確認し、自身の望む画像を捜索しなければならない。その結果、ユーザは撮影日時、キャプション、「タグ」の情報を用いた検索を効率よく行う事が不可能になっている。図 2-3 は、ウェブブラウザの Narrative Clip 閲覧システムでアルバム一覧を表示した画像である。図 2-4 は、ウェブブラウザの Narrative Clip 閲覧システムで、あるアルバムを選択し、アルバム内の画像をリスト表示している画像である。

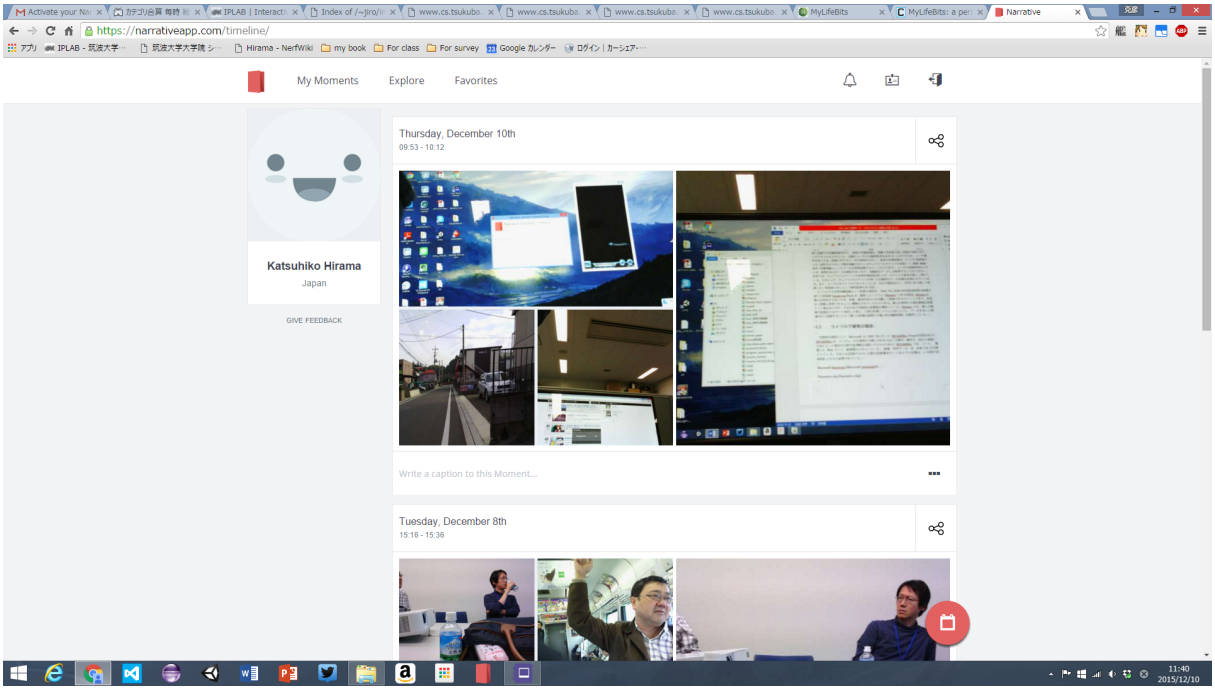



図 2-3 Narrative Clip アルバム表示

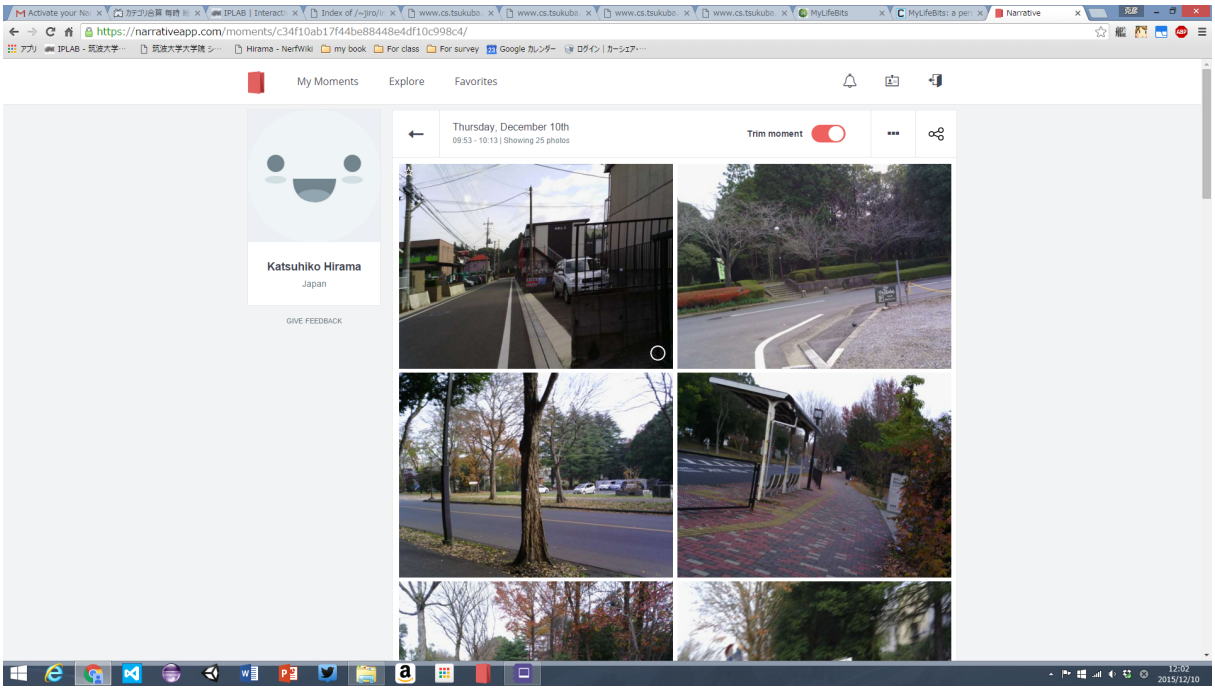



図 2-4 Narrative Clip リスト表示

# 第3章　ライフログにおける検索・閲覧の重要性

保存された情報内における検索は、ライフログ閲覧による過去の分析、追体験に重要である。仮に、ウェアラブルカメラを用いて日常生活の記録として 30 秒につき 1 枚写真を 24 時間態勢で撮影した場合、一日で 2880 枚になり、一年では約百万枚を超えることになり、ライフログは膨大な量となっていく。この中から、ユーザの求めるライフログ画像を取得することは、検索システムがなければ困難になってしまう。そのため、ライフログ閲覧には、記録された膨大な情報量を利用するためにも、ログデータを効率良く検索する仕組みが必要である。

現行の閲覧システムでは、ユーザがライフログ画像を検索・閲覧する場合、自身の記憶を頼りに、ライフログを検索しなくてはならない。このとき、ライフログ画像の検索のために、画像の時間情報(記録時刻)や、記録した画像に対してユーザが手動でメモなどを記載することで、ライフログ画像の情報を充実させている。しかし、人の記憶は非常に曖昧であるため、画像の時間情報やメモの内容を忘れてしまう可能性がある。そのため、時間やメモの情報だけでは、検索に対する情報としては不十分である。同時に、ユーザが膨大な量となるライフログ画像に対してメモを残す事は、非常に負担がかかってしまう。よって、ライフログ画像の情報を自動的に充実させる必要がある。

一般の画像検索と、ライフログ画像の検索では、異なる部分がある。前者では、物体または人物の名称・形状などの情報が確かな状態である（例:友人 A が映る写真を調べる、今撮影した物体について調べる)。加えて、検索結果画像の持つ時間情報や位置情報はあまり重要でない。後者では、検索したい事柄が不確かな状態で検索する場合がある（例:名前は知らないが、この写真の人について調べたい。あの日持っていたものを調べたい)。さらに、検索結果画像の持つ時間情報や位置情報は、ライフログとして重要である。これは、ライフログ画像の閲覧により、記憶想起が行うことが可能であることが、関係していると考える。

特に物体・人物の情報は、ライフログ検索・閲覧に重要であると考える。これは、記憶想起の一つとして、自身が保持している記憶を、物事の連想から想起する「再生」[10]にも重要とされているからである。Wagenaar が行った出来事の記憶と「再生」の関係に関する研究[11]において、6 年間に渡り 2,400 件の出来事に対して、「だれと(Who)」「なにを(What)」「どこで(Where)」「いつ(When)」について記録をした。記憶想起を行う際に、記録した手がかりのうちの 1 つを自身に示して、他の手がかりを含めて出来事を再生した。その結果、最も強力な手がかりは「なにを」であり、続いて「だれと」「どこで」となり、「いつ」は、ほとんど役に立たなくなった。Matthew ら[12]は、アルツハイマー病などにより記憶障害を持つ患者が記憶想起を行う場合、どのような情報を手掛かりとすることで、記憶想起を促せるかを述べた。また、エピソード記憶障害の有無によって、記憶想起がどのように異なっていたかを調査した。結果、物体情報(What)、人物情報(Who)、行動情報(How)が記憶想起のきっかけとして多く使われていた。これより、「いつ」ユーザが経験した日時、「どこで」ユーザがいた場所に加えて、「なにを」ある時点でユーザの周りに何が存在していたのか、という情報と、「だれと」ユーザの近くにいる人々の情報は、関連づけられた過去の情報取得に有用である。これらの情報を用いることで、ユーザはライフログ活用の一つである記憶想起を効率的行うことができる。

# 第4章　研究目的

現行の閲覧システムでは、ライフログ画像の持つ情報を用いた検索が十分に行えていないこと、ライフログ閲覧閲覧の活用である記憶想起に対してライフログ画像の持つ情報が十分に提供されていないことが問題である。

そこで本研究は､ライフログ画像の情報を用いた検索･閲覧システムの構築を目的とする。これにより、ユーザは過去の分析や追体験を可能となる。そのために、「いつ(When)」「どこで(Where)」だけでなく、記憶想起に重要である「だれと（Who)」「なにを(What)」の情報を、ライフログ画像内に映る人物・物体を検出・認識することで取得する。取得した情報から、ライフログ画像の情報を自動的に充実させる。しかし、画像情報を充実するだけでは、ユーザが人物を忘れてしまう・名前を記憶していないことや、広く意味を持つ物体について検索する際に、目的のライフログを検索する事が難しくなる問題点がある(例:「赤い鞄」について調べたいが、自動的なタグ付けでは一般化され、「鞄」となってしまう。「鞄」では、検索結果が多くなり、ユーザが求める「赤い鞄」についてのライフログ画像をさらに探さなくてはいけない)。その結果、ライフログ画像の情報を充実させても、ユーザが自身の過去の分析等が行えない場合がある。この問題点に対して、ユーザが検索したい人物・物体の画像を用いる方法を提案する。本研究では、人物・物体を含む画像を入力に用いたログデータ検索システムを開発する。そして、検索システムを導入したライフログ画像の閲覧システムを構築する。

# 第5章　想定システム

本研究では、Vicon revue を使用して、取得できるライフログ画像を用いる。これは、Vicon revue に搭載されている広角レンズにより、多くの物体・人物を写真におさめる事ができるため、使用者の経験を特徴づける物体や人物を含んだ画像を撮影する事に適しているからである。撮影された各ライフログ画像には撮影時間による時間情報と、機器に搭載されている GPS による位置情報を保持している。撮影された画像は撮影時間・撮影場所についての情報と共に、ライフログ画像用データベース内に保存する。ライフログ画像内には、人物・物体が写り込んでいるため、人物の検出・認識と物体検出・認識を行い、人物・物体情報を取得する。その結果、データベースには、以下の情報がライフログ画像一枚ずつに保存されている。

- 時間情報・・・画像の撮影時間
- 位置情報・・・外部デバイスの GPS 情報
- 人物情報・・・ライフログ画像内の人物位置、人物名など
- 物体情報・・・ライフログ画像内の物体位置、物体名など

想定システムでは、時間情報・位置情報・人物情報・物体情報を用いた検索を行う。ユーザは検索する時間範囲、場所や店などの情報、共に行動していた人物、ユーザが所持していた、または周りにあった物体情報、そして撮影した画像を入力に用いる（例:～月～日に、スーパーにいた時、この顔写真の人と出会った場所など)。ユーザはすべての情報に対して入力する必要はなく、ユーザが入力した内容によって、検索する種類が変化する。時間情報検索では、入力に該当する時間情報を持つライフログ画像が取得される。位置情報検索では、入力した位置情報や店名などから、該当する位置情報を持つライフログ画像が取得される。人物情報では、入力情報と同一人物が映るライフログ画像が取得される。物体情報では、入力情報と同一物体が映るライフログ画像が取得される。それぞれの検索によって得られた検索結果を組み合わせることで、複数のユーザ入力に対した検索結果を出力する。図 5-1 は想定システムの概要を図に示したものである。

検索結果の閲覧は、ユーザが検索結果を一目で見分けられるようにサムネイルリストで表示する。ユーザは自身が入力した画像などと同時に、検索結果を閲覧することができる。そのため、ユーザ入力を左に表示し、右に検索結果がサムネイルリストで表示されるインターフェースが想定できる。図 5-2 は検索結果閲覧のユーザインターフェースのイメージ図を示したものとなる。サムネイルリスト内の画像を選択することで、画像拡大閲覧を行うことができる。

画像拡大閲覧では、選択した画像が元サイズで表示されると同時に、画像の持つ時間情報・位置情報が表示される。画像内の人物または物体部分にマウスカーソルを合わせると、該当部分の人物・物体情報を表示することができる。この時、人物または物体を選択することで、選択した人物・物体に対しての検索を行うことができる。これにより、人物・物体情報と通して、自身の記憶やライフログを関連付けて取得・閲覧し、記憶想起をすることができる。, 図 5-3 は画像拡大認識のイメージ図を示したものである。

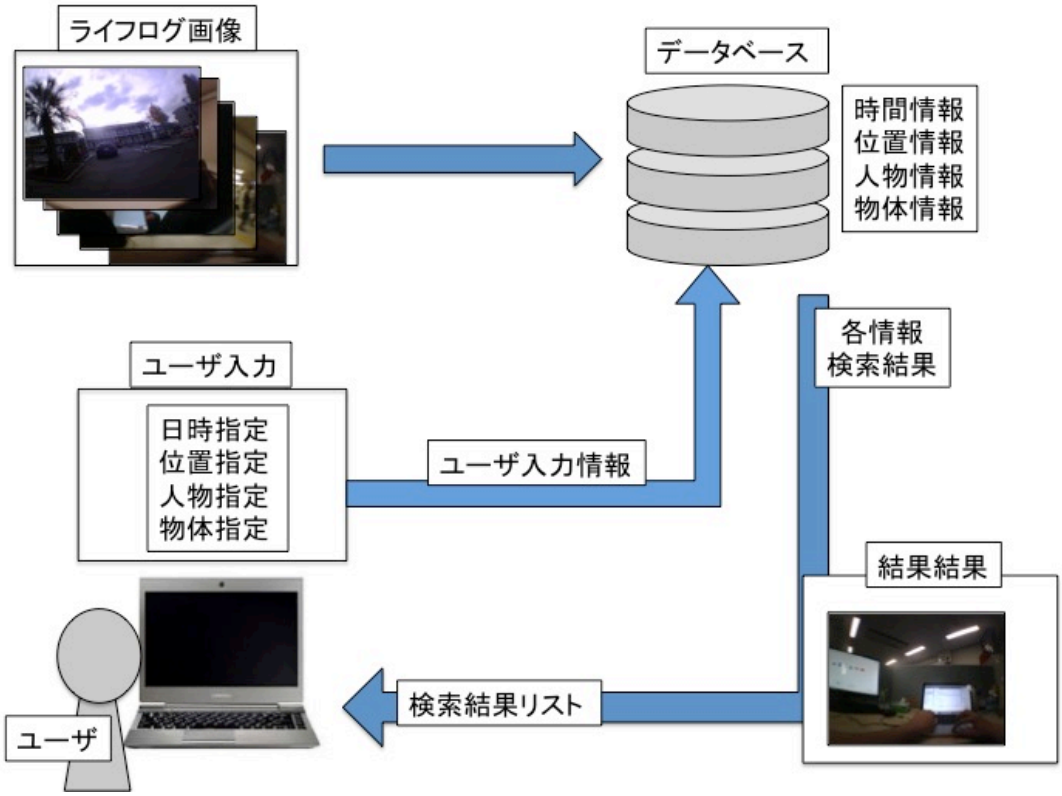



図 5-1 想定システム概要

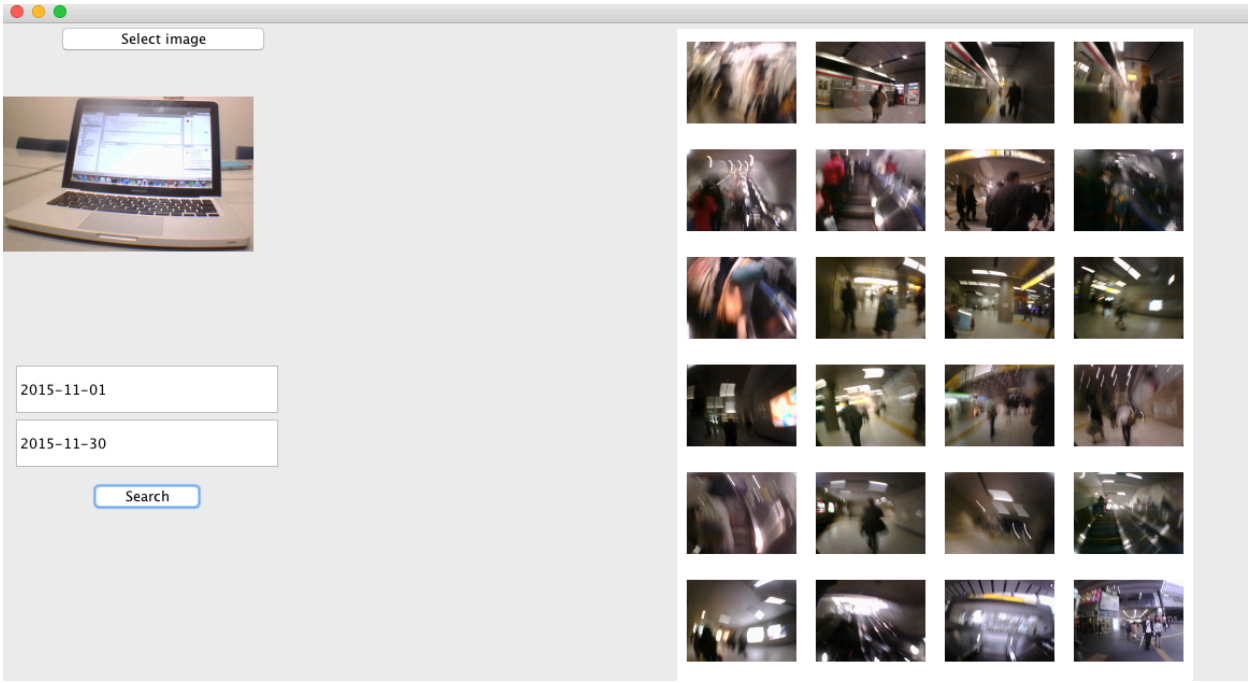



図 5-2 検索結果閲覧イメージ図

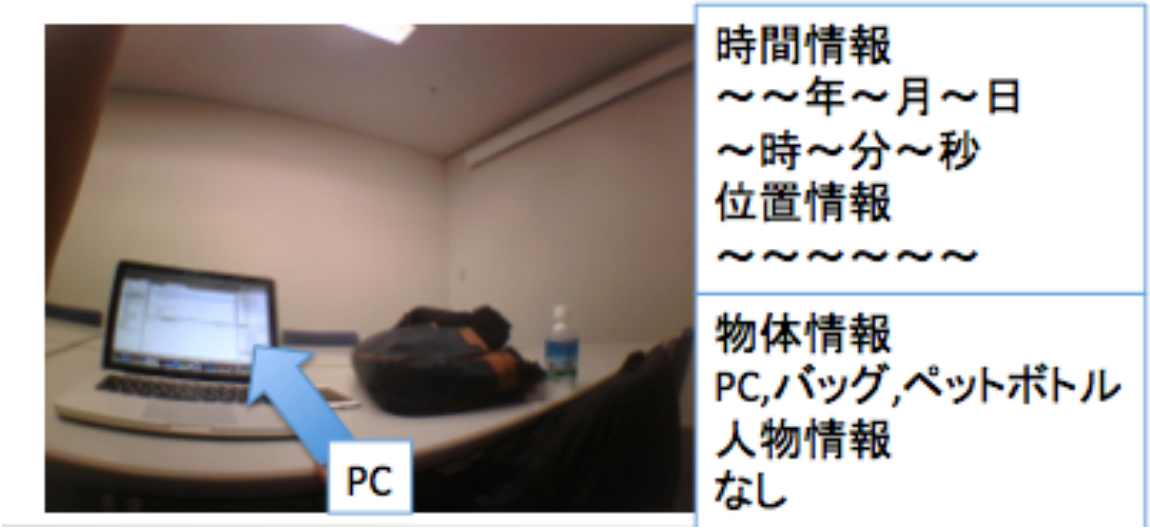



図 5-3 画像拡大認識イメージ図

# 第6章　プロトタイプシステムの部分実装

## 6.1　概要

想定システムを踏まえたプロトタイプシステムでは、データベースにライフログ画像を保存する際に、各ライフログ画像に対して顔検出と物体検出・物体名の取得を行う。顔検出では、ライフログ画像内における人物の顔位置を取得し、物体検出ではライフログ画像内における物体位置を取得する。物体名の取得では、物体検出によって取得した物体画像に対して、画像認識を行い、検出された物体の名称を取得する。これらの情報を各ライフログ画像の人物・物体情報としてデータベースに保存する。

ユーザの入力は、人物または物体が映る画像と日時・位置の指定となる。入力された日時・位置指定は、それぞれ時間情報・位置情報の検索に用いる。入力した画像は、まず顔検出・物体検出を行う。顔検出によって、画像内に人物の顔が検出された場合、画像内の顔部分を抽出し、人物情報の検索に用いる。物体検出では、検出された物体に対して、物体名取得のための画像認識を行う。物体検出から得られた物体画像と物体名を用いて、物体情報の検索を行う。人物・物体情報の検索には、画像特徴量のマッチングによって検索を行う。マッチング結果によって、同一人物または同一物体が映るライフログ画像を、ユーザは閲覧することができる。図 6-1 は、プロトタイプシステムの概要を示したものである。

今回、プロトタイプシステムの部分実装として、人物情報の取得にあたる「顔検出」、物体情報の取得にあたる「物体検出」、そして人物・物体情報の検索部分にあたる「画像特徴量マッチング」の実装を行った。使用言語は、GUI 部分に Java、検出・認識部分に Python を用いた。

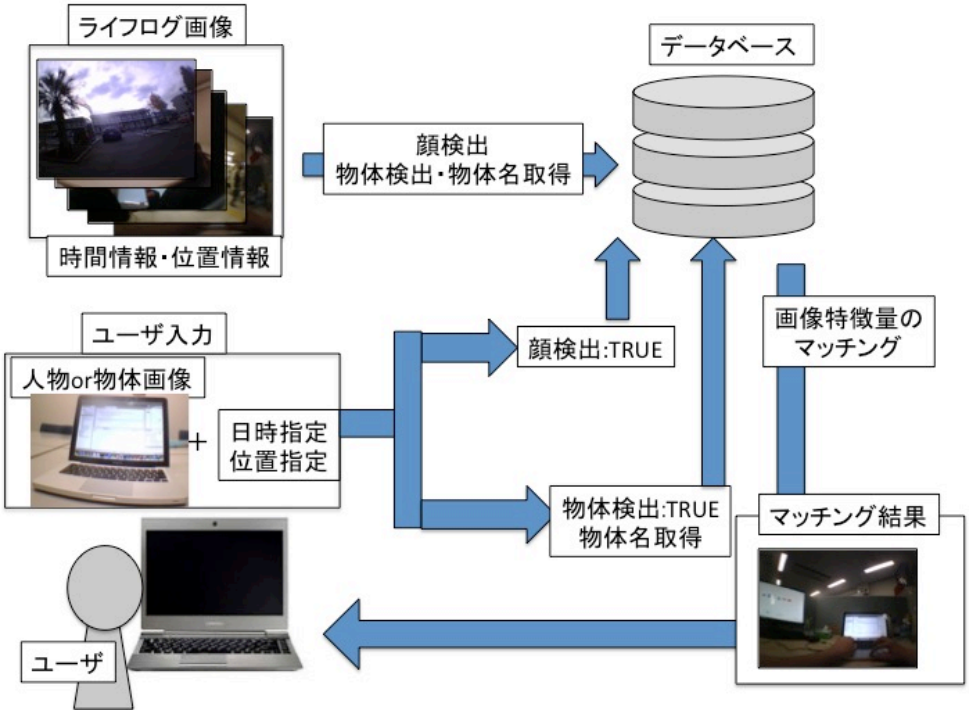



図 6-1 プロトタイプシステム概要図

## 6.2 顔検出

### 6.2.1 概要

画像内の顔検出には、OpenCV[13]にある Haar-like 特徴量とカスケード分類器を用いた顔認識を採用した。この検出アルゴリズムは、Paul ら[14] のオブジェクト検出をベースに、Rainer ら[15]の研究によって改良されたものを用いている。カスケード分類器は、OpenCV で提供されている haarcascade_frontalface_alt.xml を使用する。顔検出によって、画像中に顔が確認できた場合、そのライフログ画像に人物が含まれていることと、画像内における顔の位置をデータベースに保存する。顔検出によって画像中に顔が確認できなかった場合、画像中に人物がいないことをデータベースに保存する。図 6-2 は実際に撮影されたライフログ画像、図 6-3 は、図 6-2 のライフログ画像に顔検出を行い、得られた顔部分を矩形で囲ったものである。

図 6-2 顔検出前ライフログ画像

図 6-3 顔検出後ライフログ画像

### 6.2.2 Haar-Like 特徴量

Haar-Like 特徴量とは、局所特徴量のうち輝度に着目したものである。局所特徴量は、抽出対象物体の局所的な領域に着目した特徴量の事であり、環境変化や物体の形状変化にロバストな特徴である。画像の明暗差により特徴を捉えており、画像の一部分を切り出し、局所的な明暗差を算出する。局所的な特徴を組み合わせることで、物体を判別可能にしている。Haar-Like 特徴量の計算では、入力画像上の任意の場所に、検出窓を配置している。検出窓中の特徴量から、検出窓内の画像を図 7-4 に示す矩形領域の組み合わせに変換することができる。図 6-4 内の矩形領域の白黒パターンを Haar-Like 特徴と呼ぶ。検出窓内の矩形位置と特徴パターンの組み合わせを用いることで、顔などのオブジェクト検出を可能にしている。Haar-Like 特徴量は、式(1)によって求める。白(r1)と黒(r2)の矩形領域の輝度平均の差を算出し、特徴量としている。S(r)は、矩形領域 r の輝度平均値を算出する関数である。

$$H(r1, r2) = S(r1) - S(r2) \tag{1}$$

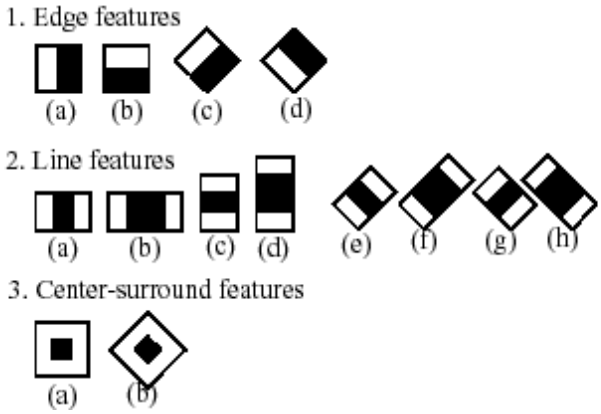



図 6-4 Haar-Like 矩形パターン例

### 6.2.3 カスケード分類器

画像認識は、学習と認識の二つのフェーズがある。学習のフェーズでは、認識したい対象画像の特徴量を抽出し、抽出した特徴量を学習データとして蓄積する。蓄積された学習データは認識のフェーズで用いられる。この蓄積された学習データが分類器となる。

カスケード分類器は、複数の分類器が一列につなげられたものである。分類の際、先頭の分類器から順番に適用していき、顔であるかの判定を行っていく。各分類器において、顔判定されれば次の分類器に処理が移り、途中で顔でないと判定された時点で、それ以降強識別器による処理が破棄される。前半の分類器では比較的少数の特徴だけを使って明らかに顔ではない分類器を棄却する。後半では、よりたくさんの特徴を使って複雑な分類を行うようにすることで、全体として特徴を計算する処理を大幅に削減するができる。図 7-4 は、カスケード分類器のアルゴリズムを示したものである。

カスケードの各段において，学習用顔画像を通過させる割合(学習用顔画像通過率）を $D_r(0 < D_r( < 1)$，学習用非顔画像を通過させる割合（学習用非顔画像通過率）を $F_p$（$0 < F_p$

< 1)とすると，カスケード n 段通過後は、学習用顔画像は $D_r^n$ 、学習用非顔画像は $F_p^n$だけ通過していることになる。

カスケードの各段を構成する識別器 $H_t(x)$ は, 図 6-5 のような矩形特徴を用いて評価を行う多数の弱識別器$h_t(x)$の線形結合により、式(2)のように表される。

$$H_i(x) = \left( sgn \sum_{t=1}^{T} \alpha_t \, h_t(x) \right) \tag{2}$$

ここで、T は用いられる分類器の数であり、$\alpha_t$は学習時に決まる分類器のエラー$\epsilon_t$を用いて以下のようにと表される。

$$\alpha_t = \log \frac{1 - \epsilon_t}{\epsilon_t} \tag{3}$$

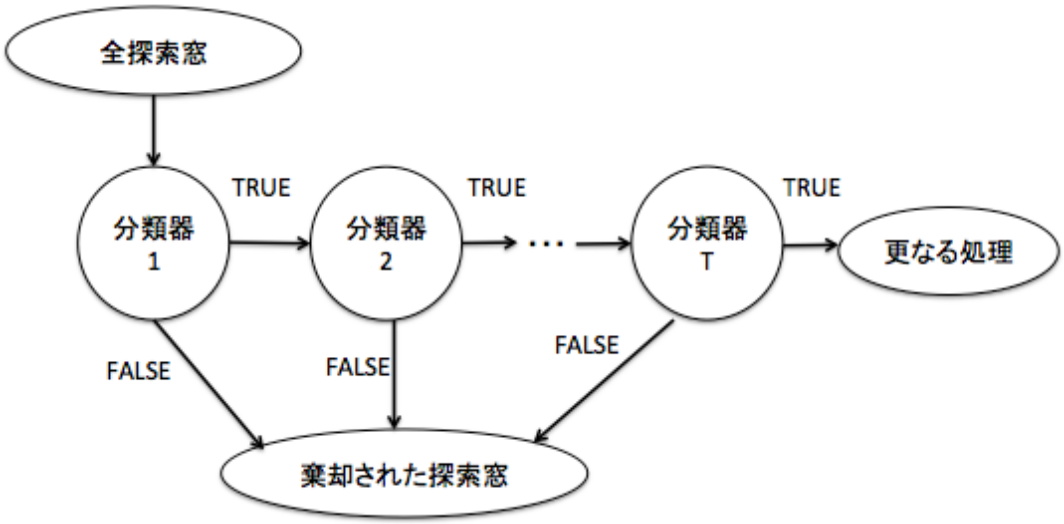



図 6-5 カスケード分類器アルゴリズム

## 6.3 物体検出

### 6.3.1 概要

ライフログ画像には複数の物体が含まれている事が想定される。そこで、Localization と呼ばれる、複数の物体が写っている可能性を考慮した物体認識を行う必要がある。Localization では、画像内の複数物体を個別に認識し、画像中での位置（矩形）の絞り込みを行う。Localization を実行する方法として、Uijlings ら[16]の Selective Search を用いる。Selective Search では、ピクセルレベルで類似する領域をグルーピングしていき、候補領域を選出している。Selective Search により、ライフログ画像内における物体位置を取得する。取得された位置から、物体の画像を切り出し、物体名の取得を行う。物体名の取得には、Labellio[17]を用いて作成した画像モデルデータベースを用いる。画像モデルデータベースから得られた物体名と、Selective Search によって得られたライフログ画像中の物体位置をタグとして、データベースに保存する。この情報を画像認識・検索に用いる。

### 6.3.2 Selective Search

Selective Search では、Pedro ら[18]の手法を用いて、入力画像を小領域へ分割を行う。次に、小領域における特徴を求める。各小領域の特徴は、以下のものを用いている。

- 色特徴(正規化色ヒストグラム)
- テクスチャ特徴(正規化勾配ヒストグラム)
- 小領域の面積
- 小領域の外接矩形

小領域の特徴用いて、類似度を計算、数値化した混合メトリクスを用意する。類似度の高い領域同士から結合していき、最終的に一つの領域になるまで繰り返すことで、候補領域抽出を行っている。これにより、物体があると思しき部分を求めることができる。図 6-6 は実際のライフログ画像に Selevtive Search を用いた結果、物体部分を矩形で囲ったものである。

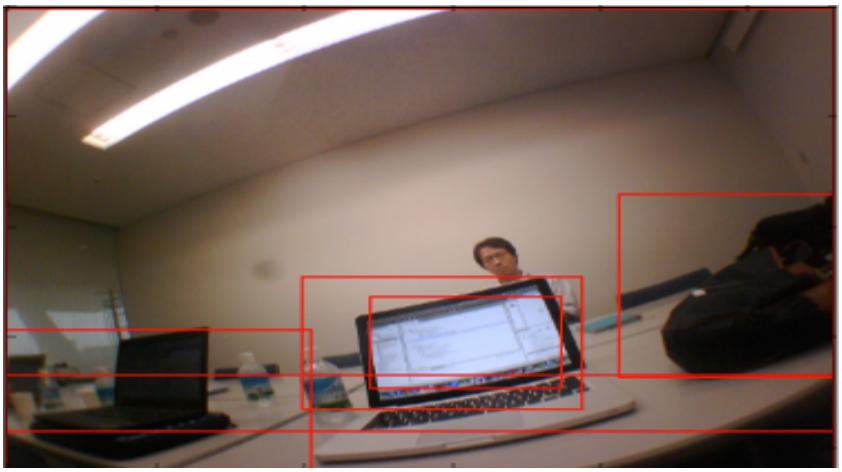
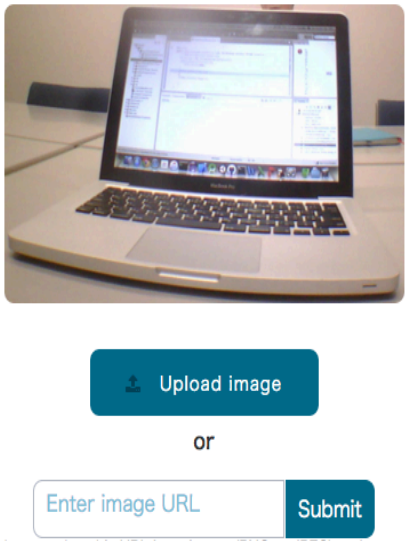


図 6-6 SelectiveSearch 処理後画像

### 6.3.3 Labellio

Labellio は、Alpaca DB. Inc.が提供している Deep Learning による画像認識モデルを自動作成するプラットフォームである。ユーザが認識したい複数の物体画像を用意し、各物体にラベル付けを行う。すべての画像に対してラベル付けを終えると、自動的に画像認識モデルを作成してくれる。この画像認識モデルは Amazon Web Service を用いて、自身の画像認識モデルを用いた web api を用意できる。これにより、Http post メソッドから画像を入力することで、最も類似度の高いラベルの情報を json 形式で得ることができる。図 6-7 は Selective Search によって得られた物体画像を、Labellio プラットフォームにおいて、自身の画像認識モデルに適応した結果となる。




| Labels | Score |
|---|---|
| パソコン | 1.00 |
| テレビ | 0.00 |
| 携帯電話 | 0.00 |
| 猫 | 0.00 |
| 犬 | 0.00 |

図 6-7 Labellio の画像認識モデル適用結果

## 6.4 画像特徴量マッチング

### 6.4.1 概要

顔検出または物体検出から得られた画像に対して、Scale-Invariant Feature Transform(以下、SIFT と呼ぶ) [19]を用いて、画像の特徴量である特徴点と特徴ベクトルを計算する。次に、OpenCV 内メソッドの DescriptorMatcher を用いて、検索入力画像の特徴量とライフログ画像の特徴量のユークリッド距離を計算する。計算されたユークリッド距離の最小値から、入力画像とライフログ画像内の人物または物体が同一であるかを判定する。

### 6.4.2 SIFT (Scale-Invariant Feature Transform)

SIFT は，特徴点(以下、キーポイントと呼ぶ)の検出（Detection）と特徴量の記述(Description)の 2 段階からなり、各処理は以下の流れとなる。

Detection
- 1.スケールとキーポイント検出
- 2.キーポイントのローカライズ

Description
- 3.オリエンテーションの算出
- 4.特徴量の記述

1. スケールとキーポイント検出は、処理によりスケールとキーポイントを検出する。2. キーポイントのローカライズは、検出されたキーポイントから特徴点として向かない点を削除し、その後サブピクセル推定を行う。3. オリエンテーションの算出は，回転に不変な特徴を得るためにキーポイントのオリエンテーションを求める。4. 特徴量の記述は、求めたオリエンテーションに基づいてキーポイントの特徴量を記述する。以下に各処理の詳細を述べる.

1. スケールとキーポイント検出

第 1 段階のキーポイント検出では、Lowe[19]によって提案された効率的な極値検出法である Difference-of-Gaussian（以下、Dog と呼ぶ）処理を用いて、スケールスペースにおける極値探索を行う。これにより、キーポイントの位置とスケールを決定する。
ガウシアンフィルタのスケールをσ、注目画素からの距離を x,y および u,y とすると、キーポイント候補点は、スケールの異なるガウス関数 G(x,y,σ)と入力画像 I(u,v)を畳み込んだ平滑化画像 L(u,v,σ)の差分(Dog 画像)から求める。それぞれ以下の式から求める。

$$L(x,y,\sigma) = G(x,y,\sigma) * I(u,v) \tag{1}$$

$$G(x,y,\sigma) = \frac{1}{2\pi\sigma^2} exp\left(-\frac{x^2+y^2}{2\sigma^2}\right) \tag{2}$$

DoG の結果の画像を D(u, v, σ)とすると、DoG 画像は次式で求まる。

$$
\begin{aligned}
D(u,v,\sigma) &= \big(G(x,y,k\sigma) - G(x,y,\sigma)\big) \\
&\qquad * I(u,v) \\
&= L(u,v,k\sigma) - L(u,v,\sigma)
\end{aligned} \tag{3}
$$

この処理を、$\sigma_0$から$k$倍ずつ大きくした異なるスケール間で行い、複数の DoG 画像を求める。$\sigma$が一定の割合で増加し続けると、ガウシアンフィルタのウィンドウサイズが大きくなり、処理できない端領域の拡大と計算コストの増加という問題が発生する。この問題に対し、画像のダウンサンプリングにより$\sigma$の変化の連続性を保持した平滑化処理を実現している。

$\sigma$の連続性を保持した効率的な平滑化処理は、はじめに，入力画像を初期値である$\sigma_0$で平滑化を行い，平滑化画像 L1($\sigma_0$) を得る。次に、$\sigma_0$を k 倍した値 $k\sigma_0$で平滑化を行い L1($k\sigma_0$) を得る。同様の処理により、$\sigma$の異なる複数の平滑化画像を得る。ここまでの処理を 1 オクターブとする。複数生成された平滑化画像の中から$2\sigma_0$で平滑化された画像$L_1(2\sigma_0)$を 1/2 のサイズにダウンサンプリングする。1 オクターブにおける処理回数については、増加率 k の設定とともに後述する。1/2 のサイズにダウンサンプリングされた画像$L_2(\sigma_0)$と、$2\sigma_0$で平滑化を行った画像$L_1(2\sigma_0)$には以下のような関係が成り立つ。

$$
L_1(2\sigma_0) \approx L_1(\sigma_0) \tag{4}
$$

この関係を利用することで、$\sigma$の最大値を制限することができるため、ガウシアンフィルタのウィンドウサイズによる計算量の増加を防ぐことができる。

$\sigma$の増加率 k は、1 オクターブにおけるスケールスペースの分割数により決定する。スケールスペースの分割数を s とした場合、1 オクターブでは、スケールスペースは$\sigma_0$から$2\sigma_0$まで増加するため、$\sigma$の増加率 k は$k = 2^{1/s}$となる。極値探索には DoG 画像を 3 枚 1 組で処理する必要があるため、s 枚の極値検出の対象となる画像を得るためには$s + 2$枚の DoG 画像が必要となる。さらに、$s + 2$枚の DoG 画像を得るためには$s + 3$枚の平滑化画像が必要になる。したがって、1 オクターブにおける平滑化の回数は$s + 3$回となる。

DoG は異なるスケールによる平滑化画像の差分のため、DoG の値が大きくなる$\sigma$では，スケールの変化領域にエッジ等の情報量を多く含んでいるといえる。そこで、DoG 画像から極値を検出し、キーポイントとスケールを決定する。極値の検出は DoG 画像 3 枚一組で行う。DoG 画像の注目画素と、その周りの 26 近傍を比較する。極値であった場合、その画素をキーポイント候補点として検出する。このような極値検出は、$\sigma$の値の小さい DoG 画像から行う。一度極値が検出された画素は、より大きなスケールで極値が検出されてもキーポイント候補点としない。この処理をスケールの異なる DoG 画像の全画素に対して行う。スケールスペースの極値の性質として、画像サイズが 2 倍になると、DoG の極値探索により検出されたキーポイントのスケール$\sigma$も比例して 2 倍となる。SIFT は、特徴を最も含むスケール$\sigma$を自動的に決定するため、空間的に同範囲の領域から特徴量を記述することで、拡大・縮小に不変な特徴量となる。

## 2. キーポイントのローカライズ

検出されたキーポイント候補点の中には、DoG 出力値が小さい点(low contrast)やエッジ上の点が含まれており、これらの点はノイズや開口問題に影響を受け易いという問題がある。そこで、キーポイント候補点の中から、主曲率とコントラストにより安定したキーポイントに絞り込む。さらに、キーポイントのサブピクセル推定により位置とスケールを算出する。エッジ上に存在するキーポイント候補点の削除方法は、まずキーポイント候補点における 2 次元ヘッセ行列 H を次式により計算し、主曲率を求める。

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix} \tag{5}$$

行列内の導関数は、キーポイント候補位置での DoG 出力値の 2 次微分から得られる。ここで、ヘッセ行列から求められる第 1 固有値を $\alpha$，第 2 固有値を $\beta$ ($\alpha > \beta$)とする。このときヘッセ行列の対角成分の和$T_r(H)$と行列式$Det(H)$は次のように計算できる。

$$T_r(H) = D_{xx} + D_{yy} = \alpha + \beta \tag{6}$$

$$Det(H) = D_{xx}D_{yy} - \left(D_{xy}\right)^2 = \alpha\beta \tag{7}$$

さらに、$\gamma$を第 1 固有値と第 2 固有値の比率とし、$\alpha = \gamma\beta$とすると次式のようになる。

$$\frac{T_r(H)^2}{Det(H)} = \frac{(\alpha+\beta)^2}{\alpha\beta} = \frac{(\gamma\beta+\beta)^2}{\gamma\beta^2} = \frac{(1+\gamma)^2}{\gamma} \tag{8}$$

この値は固有値そのものではなく、固有値$\alpha, \beta$の比率で決まる。したがって、固有値を求めずにエッジ上の点であるか判別することが可能となる。この値を次式に示すようにしきい値処理することで、不要なキーポイント候補点を削除する。

$$\frac{T_r(H)^2}{Det(H)} < \frac{(1+\gamma_{th})^2}{\gamma_{th}} \tag{9}$$

式()を満たす点をキーポイント候補する。閾値は、$\gamma_{th}$により決定する。この処理により、固有値の比率が閾値より大きい点、つまりエッジ上に存在する点が削除される。

3 変数$(x, y, \sigma)$の 2 次関数をフィッティングすることで、キーポイント候補点のサブピクセル位置とスケールを算出する。ある点$x = (x, y, \sigma)^T$ での DoG 関数 D(x) をテイラー展開すると次式のようになる。

$$D(x) = D + \frac{\partial D^T}{\partial x}x + \frac{1}{2}x^T\frac{\partial^2 D}{\partial x^2}x \tag{10}$$

上式について、x に関する偏導関数を求め、0 とすると次式が得られる。

$$\frac{\partial D}{\partial x}+\frac{\partial^2 D}{\partial x^2}\hat{x}=0 \tag{11}$$

このとき$\hat{x}$はキーポイント候補点（極値）のサブピクセル位置を表している。この式を変形し次式を得る。

$$\frac{\partial^2 D}{\partial x^2}\hat{x}=-\frac{\partial D}{\partial x} \tag{12}$$

上式は以下のように表される。

$$\begin{bmatrix}\frac{\partial^2 D}{\partial x^2} & \frac{\partial^2 D}{\partial xy} & \frac{\partial^2 D}{\partial x\sigma}\\ \frac{\partial^2 D}{\partial xy} & \frac{\partial^2 D}{\partial y^2} & \frac{\partial^2 D}{\partial y\sigma}\\ \frac{\partial^2 D}{\partial x\sigma} & \frac{\partial^2 D}{\partial y\sigma} & \frac{\partial^2 D}{\partial \sigma^2}\end{bmatrix}\begin{bmatrix}x\\ y\\ \sigma\end{bmatrix}=-\begin{bmatrix}\frac{\partial D}{\partial x}\\ \frac{\partial D}{\partial y}\\ \frac{\partial D}{\partial \sigma}\end{bmatrix} \tag{13}$$

式(13)をキーポイント候補点のサブピクセル位置$\hat{x}$を得るために、変形する。

$$\begin{bmatrix}x\\ y\\ \sigma\end{bmatrix}=-\begin{bmatrix}\frac{\partial^2 D}{\partial x^2} & \frac{\partial^2 D}{\partial xy} & \frac{\partial^2 D}{\partial x\sigma}\\ \frac{\partial^2 D}{\partial xy} & \frac{\partial^2 D}{\partial y^2} & \frac{\partial^2 D}{\partial y\sigma}\\ \frac{\partial^2 D}{\partial x\sigma} & \frac{\partial^2 D}{\partial y\sigma} & \frac{\partial^2 D}{\partial \sigma^2}\end{bmatrix}^{-1}\begin{bmatrix}\frac{\partial D}{\partial x}\\ \frac{\partial D}{\partial y}\\ \frac{\partial D}{\partial \sigma}\end{bmatrix} \tag{14}$$

上式を解くことにより、キーポイント候補点のサブピクセル$\hat{x}=(x,y,\sigma)$を得る。

サブピクセルの位置での Dog 出力を算出し、コントラストによるキーポイントの絞り込みを行う。式()は次のように表される。

$$\hat{x}=-\frac{\partial^2 D^{-1}}{\partial x^2}\frac{\partial D}{\partial x} \tag{15}$$

上式に式(10)を代入すると次式が得られる。

$$D(\hat{x})=D+\frac{1}{2}\frac{\partial D^T}{\partial x}\hat{x} \tag{16}$$

D は DoG 関数であり、x^ はサブピクセル位置を表しているため、上式はサブピクセル位置での DoG 出力値となる。この DoG の値からキーポイント削除の判別を行う。

## 3. オリエンテーションの算出

検出された各キーポイントのオリエンテーションを求める。オリエンテーションはキーポイントにおける方向を表し、特徴量記述の際にオリエンテーションにより向きの正規化を行うことで、回転に不変となる。キーポイントのオリエンテーションを求めるには、キーポイントが検出された平滑化画像 L(u,v)の勾配強度 m(u,v)と勾配方向 θ (u,v)を以下の式により求める。

$$m(u,v) = \sqrt{f_u(u,v)^2 + f_v(u,v)^2} \tag{17}$$

$$\theta(u,v) = \tan^{-1}\frac{f_v(u,v)}{f_u(u,v)} \tag{18}$$

$$f_u(u,v) = L(u+1,v) - L(u-1,v) \tag{19}$$

$$f_v(u,v) = L(u,v+1) - L(u,v-1) \tag{20}$$

局所領域における勾配強度 m(x, y)と勾配方向 θ (x, y) から重み付方向ヒストグラム h を以下の式により作成する。

$$h_{\theta'} = \sum_x \sum_y w(x,y) * \delta[\theta', \theta(x,y)] \tag{21}$$

$$w(x,y) = G(x,y,\sigma) * m(x,y) \tag{22}$$

ここで、$h_\theta$は，全方向を 36 方向に量子化したヒストグラムである。w(x, y) はある局所領域の画素(x, y) での重みであり、キーポイントが持つスケールサイズのガウス窓$G(x,y,\sigma)$と勾配強度 m(x, y)から求める。δ は Kronecker のデルタ関数で、勾配方向 θ (x, y) が量子化した方向$\theta_0$に含まれるとき 1 を返す。また、このときのガウス窓にはキーポイントが持つスケールを用いる。ガウス窓による重み付けにより、キーポイントに近い特徴量がより強く反映される。この 36 方向のヒストグラムの最大値から 80% 以上となるピークをキーポイントのオリエンテーションとして割り当てる。

## 4. 特徴量の記述

検出したオリエンテーションを基に、128 次元の特徴量を記述する。まず、キーポイントのオリエンテーション方向に回転する。特徴量の記述には、キーポイント周辺領域のもつ勾配情報を用いる。使用する勾配情報は、キーポイントを中心とし、そのキーポイントが持つスケールを半径とした円領域内から求める。周辺領域を、一辺 4 ブロックの計 16 ブロックに分割し、ブロックごとに 8 方向(45 度ずつ)の勾配方向ヒストグラムを作成する。この勾配方向ヒストグラムは、キーポイントのオリエンテーションを算出したときに作成したヒストグラムと同様の手法で求める。キーポイントが持つオリエンテーション方向に座標軸をあわせた領域で特徴を記述するため、回転に不変な特徴量となる。また，128 次元の各特徴ベクトルの長さはベクトルの総和で正規化する。これにより、キーポイントは照明変化に対して影響の少ない特徴量となる。

### 6.4.3 対応点探索による画像マッチング

SIFT により異なる画像間で抽出された各キーポイントの特徴量を比較することで、画像間の対応点探索が可能となる。以下に、対応点探索の流れを示す。

画像 I1 中にあるキーポイント$k_{I1}$と画像 I2 中にあるキーポイント$k_{I2}$の特徴量をそれぞれ、$v^{k_{I1}}$、$v^{k_{I2}}$とすると、特徴量間のユークリッド距離 d は次式により算出される。

$$d\left(v^{k_{I1}}, v^{k_{I2}}\right) = \sqrt{\sum_{i=1}^{128}\left(v_i^{k_{I1}} - v_i^{k_{I2}}\right)^2} \tag{23}$$

ここで、SIFT 特徴の次元数は 128 次元である。あるキーポイント 1 点に対して、異なる画像中に含まれる全キーポイントとの特徴量間の距離 d を算出し、その中で最も d が最小となる点同士を対応点として検出する。SIFT 特徴量を用いると、スケール変化に影響を受けず、対応点の検出が可能であり、対応点探索を行うことで、物体認識が可能である。

# 第7章　関連研究

## 7.1　ライフログの検索・閲覧に関する研究

Memon ら[20][21]は、写真ベースのログデータ効率化のために、ユーザのコンテキストに基づくログデータを取得し、検索するライフログシステムの提案、構築を行った。個人を識別するために赤外線送受信機を取り付けたライフログデバイスを作成した。ユーザ同士が会った場合に互いの赤外線コードを交換する事で、ライフログに誰と行動を行ったか、を記録できるようにした。同時に、ユーザが所持していた物体を、ライフログデバイスによって撮影された画像と画像認識 MoodStock API を用いて、システムが物体認識する。これらをログデータ検索のためのデータとして使用した。劉ら[22]により、Vicon Revue の検索・閲覧システムと Memon らのシステムを用いて比較実験が行われた。結果、Memon らのライフログシステムのログデータ検索・閲覧効率は、「Vicon Revue」より優れている事が判明した。

Akhil ら[23]は、ユーザが記述した SNS の内容や、作成したスマートフォンアプリケーションから出来事の状況や人物、音声などの情報をデータとして記録する。得られた情報から、状況を簡単に推測できるイメージ画像を生成する。このイメージ画像を用いる事で、ユーザに効率的に記憶想起を促すと同時に、第三者がユーザの追体験を行う際に有用であることがわかった。このイメージ画像があることで、ユーザが他者とライフログを共有しようとするモチベーションが上がる事もわかった。

Doherty ら[24]は、SenseCam によって得られたライフログ画像内には、約 8000 以上のイベントについての情報が保持されている、と考えた。イベントを判別するために、MPEG-7、SIFT、SURF に基づいた画像処理システムを考案した。これは「公園を歩いていた他の日」など、イベントに基づいた分析を行う事で、似たイベントを検索する際に役立つと考えたからである。評価結果より、MPEG-7 と SIFT を融合したアプローチと SIFT 個別に使用する事で効果がある事を発見した。同時に、ライフログイベントは検索にも大きな影響があることを示した。

Lijuan ら[25]は、スマートフォンなど、センサ類が充実しているモバイルデバイスを用いて、行動のタグを自動的に取得し記録するアプリケーションの提案をした。記録できたタグ毎にクラスタリングを行い、異なる活動毎に個人の生活を検索・閲覧する際に使用する。

Matthew ら[26]は、画像・音声・位置情報をログデータとして取得した。システムは、画像内に存在する記憶想起の手がかりとなる可能性があるオブジェクトを抽出する。ユーザはそれらを確認し、どのオブジェクトが記憶想起に有用であるかを選択し保存する事ができる。同時に、画像に映る状況について自身で説明した音声を、ログデータと一緒に保存することができる。このシステムを 80 歳前後の記憶障害または軽度認識障害を持つ患者に使用し、評価実験を行った。結果、一ヶ月後の記憶想起・保持が、既存の手法よりも良い状態である事が判明した。

Chen ら[27]は、イベントやアイテムが記憶想起に重要であると考えた。そこで、デジタルデバイスでの活動内容、デジタルデバイス上の活動と関連のあるファイル、SenseCam などで撮影された画像を取得した。それらの情報を時間毎に検索・閲覧できるアプリケーション”iCLIPS”を作成した。これにより、ユーザそれぞれが重要と感じるデジタル上のイベントやアイテムと実世界行動が関連づけられたログデータを検索・閲覧することができる。

## 7.2 記憶想起に関する研究

Abigail ら[28]は、過去や個人的な出来事についての記憶をサポートするために日常生活をキャプチャするライフログ技術である SenseCam の実用性評価を行った。その結果、Sensecam によって撮影された画像からユーザは、自身の過去を想起するための要因が含まれていることがわかった。しかし、記憶想起を行うためには、ユーザ自身が想起する内容を覚えている状態と知っている状態で、目的のライフログ画像を検索するための時間や方法が違うことがわかった。また、自動的に撮影した画像とユーザによって撮影された画像の有効性を比較し、今後のライフログ技術についての主張を述べた。

Brown ら[29]は、コンテキストアウェアアプリケーションの設計についての研究を行った。コンテキストアウェアとは、ユーザの興味関心などから希望する情報などを予測することである。ユーザの状況から出来事などのコンテキストの取得に関する研究も行っている。この際に、ユーザの状況などを把握する場合のオブジェクトのコンテキストの重要性が述べられている。

Cosley ら[30]は、人との出来事の回想を目的とした「Pensive」と呼ばれるシステムを提案した。Cosley らは、撮影された画像にアクセスするための情報として、ユーザ自身の記憶に頼るシステムが非常に多いとした。Pensive では、ソーシャルメディアサービスの持つ人物情報を再利用し、人物情報を利用可能にした。プロトタイプを用いた評価から主観評価などを行い、システムの改良をした。インタビューや経験から、ユーザが特定の人との出来事の想起を支援するツールの設計には、想起を支援するツール自体が簡潔であることが重要である、と述べている。

# 第8章　まとめ

本論文では、ライフログの定義を確認し、現行のライフログ収集デバイスを調査した。調査結果から、現在のライフログ画像閲覧・検索システムでは、ライフログ画像の持つ情報を用いた検索が十分に行えないことや、閲覧時におけるライフログ画像の持つ情報が十分に提供されていないことを示した。

そこで、本研究ではまず、ライフログにおける検索・閲覧の重要性と記憶想起における人物・物体情報の有用性を示した。それらを踏まえ、ライフログ画像の情報を用いた検索・閲覧システムの構築を目的とした。想定システムとして、時間・位置情報だけでなく、人物・物体情報を検索・閲覧に用いたライフログ検索・閲覧システムを提案した。想定システムを踏まえたプロトタイプシステムでは、顔検出・物体検出による人物・物体情報の取得、ユーザ入力に画像を用いた画像特徴量マッチングによる人物・物体検索を提案した。プロトタイプシステムの部分実装として、顔検出・物体検出・画像特徴量マッチングの実装を行った。

# 謝辞

本論文を執筆するにあたって，指導教員である田中二郎先生をはじめ、三末和男先生、高橋伸先生、志築文太郎先生および嵯峨智先生には、ゼミや面談などを通して、丁寧な御指導と貴重な御意見を頂きました。研究についてのみならず、勉学への姿勢などについても御指導いただき、心より感謝申し上げます。

また、インタラクティブプログラミング研究室の皆様には、研究生活全体にわたって数多くの御指摘や御意見を頂きました。特に、NREF チームの皆様には、研究室での活動を通して、多くの助力や励ましのお言葉をいただき、心より感謝いたします。

# 参考文献


[1] 総務省 平成 26 年版情報通信白書 ウェアラブル端末
http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h26/html/nc141330.html

[2] 新保 史生. “ライフログの定義と法的責任 個人の行動履歴を営利目的で利用することの妥当性”. 情報管理 53(6), pp.295-310, 2010.

[3] Vannervar Bush. “As We May Think”. The Atlantic Monthly pp112 -124 July 1945

[4] Michael K. Buckland. “Emanuel Goldberg, Electronic Document Retrieval, And Vannevar Bush's Memex”. Journal of the American Society for Information Science 43, no. 4 pp.284–294, May 1992.

[5] Jim Gemmell, Gordon Bell and Roger Lueder. “MyLifeBits: a personal database for everything”. Communications of the ACM - Personal information management Volume49 Issue 1, pp.88-95, January 2006.

[6] Jim Gemmell, Gordon Bell, Roger Lueder, Steven Drucker and Curtis Wong. “MyLifeBits:fulfilling the Memex vision” MULTIMEDIA '02 Proceedings of the tenth ACM international conference on Multimedia pp.235-238, 2002.

[7] Steve Hodges, Lyndsay Williams, Emma Berry, Shahram Izadi, James Srinivasan, Alex Butler, Gavin Smyth, Narinder Kapur and Ken Wood. “SenseCam: a restrospective memory aid” UbiComp'06 Proceedings of the 8th international conference on Ubiquitous Computing pp.177-193, 2006.

[8] Narrative “Narrative Clip” http://getnarrative.com/

[9] AutoGrapher http://www.autographer.com/#home

[10] 月浦 崇. “想起・誤想起(記憶)” 脳科学辞典
https://bsd.neuroinf.jp/wiki/想起・誤想起（記憶）

[11] William Albert Wagenaar. “My memory: A study of autobiographical memory over six years, Cognitive Psychology” Vol.18. pp.225-252, 1986-04.

[12] Matthew L. Lee and Anind K. Dey. “Providing good memory cues for people with episodic memory impairment” Assets '07 Proceedings of the 9th international ACM SIGACCESS conference on Computers and accessibility pp.131-138, 2007.

[13] OpenCV http://opencv.org/

[14] Paul Viola and Michael J. Jones. “Rapid Object Detection using a Boosted Cascade of Simple Features”, Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on Volume:1 pp.511-518, 2001.

[15] Rainer Lienhart and Jochen Maydt. “An Extended Set of Haar-like Features for Rapid Object Detection”, IEEE ICIP 2002, Vol. 1, pp.900-903, Sep. 2002.

[16] J. R. Uijlings, K. E. Sande, T. Gevers and A. W. Smeulders. “Selective Search for Object Recognition” International Journal of Computer Vision archive Volume 104 Issue 2, pp.154-171, September 2013.

[17] Labellio https://www.labell.io/

[18] Perdo F. Feizenszwalb and Daniel P. Huttenlocher. “Efficient Graph-Based Image Segmentation” International Journal of Computer Vision Volume 59 Issue 2, pp. 167-181, September 2004.

[19] D. G. Lowe. “Object recognition from local scale invariant features”, Proc. of IEEE International Conference on Computer Vision (ICCV), pp.1150-1157, 1999.

[20] Moshin Ali Memon and Jiro Tanaka. “Sharing Life Experiences with Friends Based on Individual’s Locality”, Proceedings of 15th International Conference on Human-Computer Interaction (HCI International 2013), Design, User Experience, and Usability: Web, Mobile and Product Design (Part IV), LNCS 8015, Las Vegas, NV, USA, pp.706-713, July 2013.

[21] Moshin Ali Memon, Jiro Tanaka and Tomonari Kamba. “Restrain from Pervasive Logging Employing Geo-Temporal Policies”, Proceedings of the 10th Asia-Pacific Conference on Computer- Human Interaction (APCHI2012), Matsue, Shimane, Japan, pp.201-208, August 28-31 2012.

[22] 劉曉萌 “ライフログデバイスにおける行動履歴の検索とその評価” 筑波大学院博士前期課程　修士論文 2013.

[23] Akhil Mathur, Anirban Majumder, Samik Datta, Sreedal Menon, Shipra Malhotra and Ankur Dahiya. “LifeView: a lifelog visualization tool for supporting sentimental recall and sharing” OzCHI '12 Proceedings of the 24th Australian Computer-Human Interaction Conference pp.371-380, 2013.

[24] Aiden R. Doherty, Ciarán Ó Conaire, Michael Blighe, Dublin, Alan F. Smeaton and Noel E. O'Connor. “Combining image descriptors to effectively retrieve events from visual lifelogs” MIR '08 Proceedings of the 1st ACM international conference on Multimedia information retrieval pp. 10-17, 2008.

[25] Lijuan Marissa Zhou, Cathal Gurrin and Zhengwei Qiu. “ZhiWo: activity tagging and recognition system for personal lifelogs” ICMR '13 Proceedings of the 3rd ACM conference on International conference on multimedia retrieval pp.321-322, 2013.

[26] Matthew L. Lee and Anind K. Dey. “Lifelogging memory appliance for people with episodic memory impairment” UbiComp '08 Proceedings of the 10th international conference on Ubiquitous computing pp.44-53, 2008.

[27] Yi Chen and Gareth J.F. Jones. “An Event-Based Interface to Support Personal Lifelog Search” In: HCI International 2009, San Diego, U.S.A. July 2009.

[28] Abigail J. Sellen, Andrew Fogg, Mike Aitken, Steve Hodges, Carsten Rother and Ken Wood “Do life-logging tech support memory for the past?: an experimental study using sensecam” CHI '07 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems pp.81-90, 2007

[29] P.J. Brown, J.D. Bovey and Xian Chen. “Context-aware applications: from the laboratory to the marketplace” Personal Communications, IEEE Volume: 4, Issue: 5 pp.58 -64, Oct. 1997

[30] Dan Cosley, Kathy Akey, Brian Alson, Jonathan Baxter, Mark Broomfield, Soyoung Lee and Chethan Sarabu. “Using technologies to support reminiscence.” BCS-HCI '09 Proceedings of the 23rd British HCI Group Annual Conference on People and Computers: Celebrating People and Technology pp.480-484, 2009.