

画像選別によるツーリング後の思い出振り返り支援システムの開発

杉浦 正太郎¹ 打矢 隆弘¹ 内匠 逸¹

概要：ライフログとは人の生活や行動をデジタルデータとして記録したものである。近年、ウェアラブル端末による容易なライフログの収集が可能となってきた。このようなライフログの利用法の一つとして情報の振り返りがある。その中でも画像や映像による振り返りは容易で分かりやすい。しかし、自動撮影により長時間の記録を行う場合、画像の数が膨大になり、人の手作業による分類は困難である。本研究では撮影した画像から振り返りに有用と思われる画像を抽出し、ユーザに提示するシステムを提案する。

Development of recollection support system for memories after touring by image selection

SHOTARO SUGIURA¹ TAKAHIRO UCHIYA¹ ICHI TAKUMI¹

1. はじめに

ライフログとは人の生活や行動をデジタルデータとして記録したものである。近年、スマートフォンやウェアラブル端末の普及により容易なライフログの記録が可能となってきた。ライフログの活用法としてマーケティングや健康管理などのほか、思い出の振り返りがある。思い出の振り返りの中でも映像や画像を用いた振り返りは容易で分かりやすい。振り返りに用いる映像や画像を記録するためのウェアラブル端末では、NarrativeCrip[1]やGoPro Hero5[2]など様々なライフログ用のカメラがある。これらは利用者の衣服や持ち物に装着することで操作をほとんど必要としない自動撮影が可能となる。しかし、自動撮影により長時間の記録を行う場合、映像の時間や画像の数が膨大になってしまうという問題がある。思い出の振り返りにこれらの映像や画像を用いる場合、人の手で分類や探索をする必要があり、振り返りの妨げになってしまう。本研究では振り返りの妨げとなっている分類や探索の手間を削減することを目指す。画像認識を用いて自動で膨大な画像の分類や有用な画像の推薦を行うシステムを提案する。

2. ライフログ

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負担が大きい。代表的なものとしてブログなどの日記やTwitter, FacebookなどのSNSなどがある。後者の自動記録は、GoProやAppleWatchのような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・心拍数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が、取得されるデータが限定されており、客観的なデータしか取得することができない。また常時記録するのでデータ量が膨大になり、そのままでの利用が困難である。本研究では思い出の振り返りに焦点を当て、膨大な画像データの要約を目指す。

3. 関連研究

3.1 大西らの研究

大西ら [3] は、GPS、脈拍、音声といったライフログから特徴点を抽出し、それを元に映像を要約、提示するシス

¹ 名古屋工業大学 大学院 工学研究科

テムを作成し、評価している。実験ではイベントを「休日を想定した2人組で行う2時間程度の移動を伴う行動」と定義している。被験者には2人組でこのイベントを行ってもらい、映像、GPS、脈拍、音声を記録する。

イベントで記録したGPS、脈拍、音声から特徴点を抽出し、特徴点前後の映像を取得し、つなぎ合わせることで要約映像とする。それぞれ特徴点の抽出方法は以下のようになっている。

GPS

GPSはFLAGシステム[4]を用いて「滞在」と「移動」に分類する。そしてFLAGシステムによって「滞在」と判別された時間区画の開始時刻を特徴点とする。

脈拍

音声はフレーム長10秒、シフト幅5秒で切り出し、各フレームのパワー平均を求めている。それらの値も脈拍同様にピーク抽出をし、そのピークの上位20個を音声の特徴点とする。

音声

収集した脈拍データからピーク値の上位10個を選出し、脈拍の特徴点とする。ここでピークとは、データの各要素を隣接する要素と比較し、両側の要素より大きかったものとしている。

評価実験では、はじめにイベント情報のみを提示し、自分の記憶で振り返ってもらい、次に映像を見せることでその有効性を評価する。「映像要約システムを用いて振り返りを行いたいと思うか」という質問に対しては、肯定的な意見が多数見られた。しかし、「無意識下で収集された映像によって振り返りができれば写真などの自発的な記録は必要ないと思うか」という質問に対しては否定的な意見が多く見られた。この理由として、写真には自分で撮る楽しさがあるという意見が多くあった。これより手動で記録するほどではないと判断した場面や、手動で記録できなかった場面の想起に価値を見出ししていると考察されている。このことより、本研究では手動での記録ができない場面の一例としてバイクの運転中の記録を想定する。

3.2 杉山らの研究

杉山らは、ツーリング後にもバイクを楽しむための快走支援システムを作成した[5]。このシステムでは走行中に映像の記録とともに「美しい風景の場所」「危険箇所」「迷いやすい場所」といったシーンを表すタグをデバイスのボタンを押すことによって付与する。そのタグをもとにウェブアプリケーションを用いて編集し、提示することで思い出の振り返りを支援する。しかし、走行中にボタンを押す必要があり、走行の妨げとなる。また、ブレ画像まで記録されてしまうという問題と、編集に手間がかかってしまうという問題がある。本研究ではこれらの問題の解決を目指す。

4. 提案手法

本研究では、ウェアラブルカメラであるGoPro Hero5をヘルメットに装着してのタイムラプス撮影を想定する。これにより、運転者の視線に近い映像を記録する。提案システムの概要図を図1に示す。

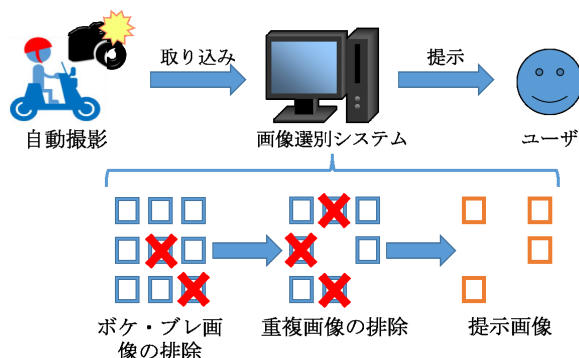


図1 提案システムの概要図

画像選別システムの手順を以下に示す。

- (1) ブレ・ボケ画像の排除
- (2) 画像の特徴抽出
- (3) 抽出した特徴から類似画像の分類
- (4) 類似画像の中から有用な画像を一枚抽出

4.1 ブレ・ボケ画像の排除

ブレ・ボケ画像を検出する方法として、エッジ検出手法を用いる。この手法では画像にガウシアンフィルタをかけ画像の平滑化をし、ラプラシアンフィルタをかけエッジ検出を行う。本システムでは以下の8近傍ラプラシアンフィルタを用いる。

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & -8 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

エッジ画像の画素値の分散が閾値より小さい場合ブレ・ボケ画像として識別する。実際の走行画像(図2)にエッジ抽出をした画像が図3となっている。



図2 元画像



図 3 エッジ画像

4.2 画像の特徴抽出

画像の特徴抽出には深層畳み込みニューラルネットワーク (DCNN) を用いる。DCNN (Deep Convolutional Neural Network) とは多層のニューラルネットワークからなる学習モデルであるディープラーニングの手法の一つであり、画像認識の分野で特に優れている。DCNN の構造の例を図 4 に示す。このように主に畳み込み層とプーリング層からなり、これらを何度か繰り返し特徴の抽出を行った後、最後に全結合層で分類を行う。

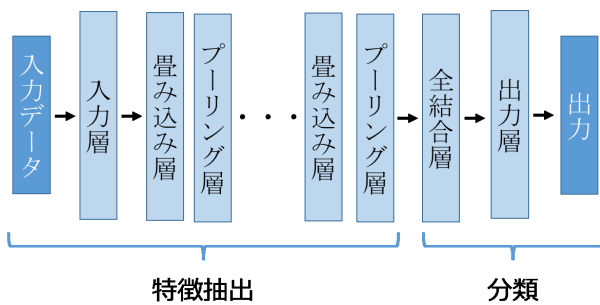


図 4 DCNN の構造例

本システムでは Google の学習済み画像識別モデル Inception-v3 の特徴抽出部分を用いる。

4.3 類似画像の分類

抽出した特徴から類似画像の分類を行う。分類には X-means 法を用いる。X-means 法とは、K-means 法によるクラスタの 2 分割を再帰的に繰り返すことにより、自動でクラスタ数を推定できるクラスタリング方式である。X-means 法の流れを以下に示す。

- (1) K(2)-means 法で 2 つのクラスタに分割
 - (2) 分割したクラスタの BIC を求める
 - (3) 分割前の BIC より分割後の BIC が小さい場合再帰的に 1 に戻る
 - (4) 分割前の BIC より分割後の BIC が大きい場合分割せずに終了
- 再帰の終了条件に用いる BIC (ベイズ情報量基準) は式 1 となっている。

$$BIC = -2 * \ln L + k \ln n \quad (1)$$

L は尤度関数、 n は標本数、 k は母数 (パラメータ数) となっている。

また入力データは DCNN の出力 2048 次元では大きすぎて計算時間が膨大になってしまうので、主成分分析を用いて 100 次元に落としてから入力する。

4.4 有用な画像の抽出

4.4.1 姫野らの研究

“イメージスケッチを用いた観光地における印象的な景観場の特性分析” という研究 [6] を紹介する。姫野らは非日常的景観における印象的な景観とはどのようなものなのかを明らかにすることを目的としている。実験方法としてインタビュー形式によるアンケート方式を実施している。印象深い景観のイメージスケッチを描かせ、非日常的景観としての印象的な景観場の抽出を行っている。また同時にそれを説明する語彙を全てイメージボキャブラリーとして抽出している。この研究では印象的な景観場の特性を明らかにするために、イメージボキャブラリーを用いたイメージ分析とイメージスケッチを用いた物理的特性分析を行なっている。イメージ分析では印象的な景観場を説明する語彙 (イメージボキャブラリー) にはどのようなものがあるのかを KJ 法を用いて明らかにする。イメージスケッチを用いた物理的特性分析ではイメージ分析で得られた印象の異なる景観群各々に、特徴的な景観構成要素や景観場の構成などを明らかにする。

分析結果では、視点、視点場と視対象、対象場の関係から、印象的な景観場に挙げられるものの 52% が遠距離景観であり、次いで中距離景観 32% が多く、近距離景観は 14% と少ないことが明らかとなった。また、視線方向は水平景が全体の 62% と最も多く、残り 38% の仰角、俯瞰景はそれぞれ同程度であることが分かった。印象的な景観場を構成する景観構成要素数は、自然物が平均 3.52 要素であるのに対し、人工物は 1.09 と低くなっている。またイメージの説明に用いられた雰囲気形容する語彙では、“広がり” “華麗” “和み” という 3 つが多く見られた。

本システムでは自然物、遠距離、広がりという要素を考慮して、空の大きな画像を印象的な画像として抽出する。

4.5 SegNet

画像中の空の領域の識別には SegNet を用いる。SegNet とはケンブリッジ大学が開発した画素単位での識別機能を実現する DeepEncoderDecoder のことである [7]。構造は図 5 のようになっている。

実際の走行画像 (図 6) に SegNet を適応したものが図 7 となっている。

色で領域の分割を行っており、それぞれ色の意味は図 8

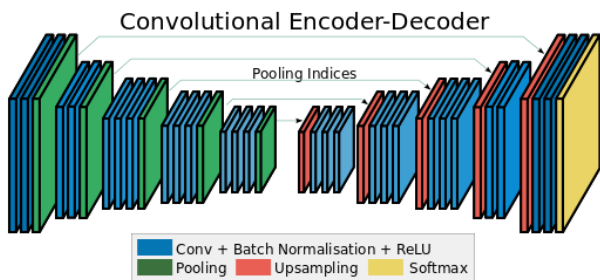


図 5 SegNet の構造



図 6 元画像

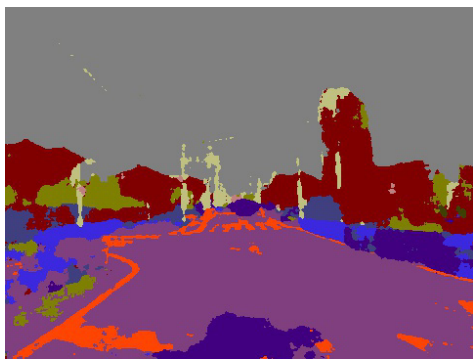


図 7 SegNet 使用後の画像

のようになっている。

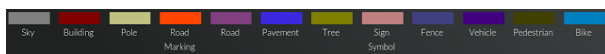


図 8 各色の意味

このうち灰色の部分が空の領域となっているので、灰色の画素数を数えることでその画像の空のサイズを計測する。各画像の空のサイズを記録し、類似画像内で一番空のサイズが大きい画像を印象的な画像として抽出する。

5. 動作実験

動作実験では被験者 1 名 (学部 4 年生) に撮影協力を依頼した。また撮影した画像をすべて利用した。

5.1 条件

動作実験の条件を表 1 に示す。

表 1 動作実験の条件

走行区間	長野県伊那郡
天候	雪→曇り→晴れ
ウェアラブルカメラ	GoPro Hero5
撮影間隔	10 秒
画像の使用枚数	642 枚 (107 分)

また今回使用した PC のスペックを表 2 に示す。

表 2 PC のスペック

機種	MacBook Air
プロセッサ	Intel Core i7 2 GHz
メモリ	8GB

処理は CPU のみとなっている。今回使用した画像の撮影位置を図 9 に示す。

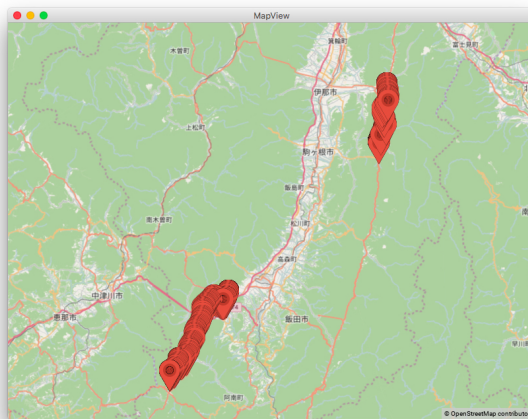


図 9 全使用画像の撮影位置

5.2 結果

抽出画像の撮影位置を図 10 に示す。

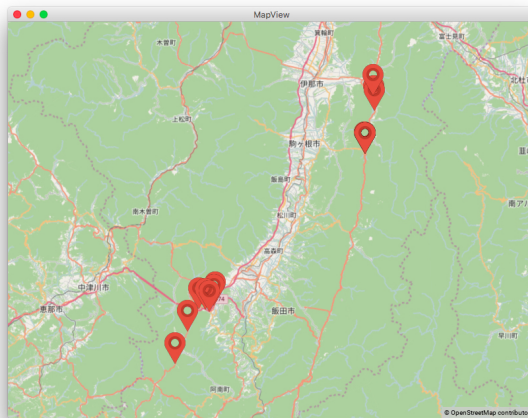


図 10 抽出画像の撮影位置



図 11 抽出画像例 1



図 12 抽出画像例 2

抽出した画像の例を図 11, 図 12 に示す。
抽出枚数と抽出処理実行時間の結果を表 3 に示す。

表 3 動作実験の結果

抽出枚数	13 枚
実行時間	73 分

5.3 考察

5.3.1 ボケ・ブレ画像の排除

GoProHero5 の手ブレ補正の精度がよく、晴れている場合にはほとんどブレやボケが発生しなかったが、雪が降っている場合ほとんどがブレ・ボケ画像となってしまった。今回は閾値を下げることで対応したが今後は適切な分類方法を検討していく必要がある。

5.3.2 類似画像の分類

今回は 107 分の画像に対し、13 のクラスに分類されている。撮影間隔を変えて行った別の動作実験では 390 枚 13 分の画像に対してもクラス数 14 と同じようなクラス数となってしまっており、撮影時間や枚数に対応できていないことがわかる。クラス数が時間や枚数あたりで一定になるように工夫していく必要がある。

5.3.3 有用な画像の抽出

基本的には空の大きな画像の抽出に成功した。しかし、

抽出画像例 2 のように地面が写った画像など空の大きな画像以外にも抽出されてしまっている。これは SegNet が地面を空と誤認識して、地面の大きな画像を抽出したことが原因だと考えられる。また、分類の時点で地面のみの画像でまとめられてしまった場合このような空の映らない画像が抽出されてしまうのだと考えられる。

6. 評価実験

6.1 実験の概要

評価実験には動作実験に用いた画像を使用する。被験者は大学生 10 名である。手順を以下に示す。

- (1) 撮影した全画像を見る
- (2) 手動で欲しいと思う画像を抽出する
- (3) 提案システムで抽出した画像を見る
- (4) 評価

質問 1. 画像による振り返りは有効か

1(有効でない)~5(有効)

質問 2. 手動での抽出に比べ手間は少ないか

1(多い)~5(少ない)

質問 3. システムで抽出した画像は振り返りに有効か

1(有効でない)~5(有効)

質問 4. 抽出枚数は適切か

1(適切でない)~5(適切)

質問 5. またシステムを用いて振り返りを行いたい

1(利用したくない)~5(利用したい)

質問 6. その他意見

6.2 結果

今回、被験者がバイクに乗るか乗らないかは考慮していない。質問 1 から質問 5 までの結果の平均と分散を表 4 に示す。質問 6 で得られた意見を以下にまとめる。

表 4 評価実験の結果 (平均・分散)

	質問 1	質問 2	質問 3	質問 4	質問 5
平均	4.1	4.6	3.5	2.6	3.9
分散	1.29	0.24	0.85	1.64	0.29

- 無駄な画像を省けているのは評価できる
- トンネルの写真が欲しかった
- 視点がずれている
- 後半の晴れてからの写真が良かったのにあまり抽出されていなかった。
- 一定時間ごとに区切って、その中から抽出して行ってはどうか
- 走行画像のみかと思っただが、駐車場での画像も案外いいなと思った
- 動画のほうが良い

6.3 考察

中間値が3なのでその値を基準に考える。

6.3.1 質問1

平均値は4.1と高くなっており、画像での振り返りが有効であることがわかる。自由記述では、どのような道を走ったか直感的に分かり嬉しい、という意見があった。

6.3.2 質問2

平均値は4.6と高くなっており、システムの利用で手間が省けていることがわかる。自由記述では、自動であるので楽になるのは当然であるという意見が多かった。また、少ないのは良いが印象に残る画像も消えてしまっていたという意見もあった。

6.3.3 質問3

平均値は3.5となっている。今回の評価実験では10人中9人は実際に走行しておらず、画像を見るのみであったので、そもそも走行画像全体が情景として印象に残らず良い評価が得られないと考えていた。しかし、結果は3.5と想定より良い結果が得られた。特に実際に走行した本人の評判が良かった。自由記述には、地面ばかりが写っているような画像も抽出されてしまっていたが、そのような画像も思い出の振り返りとしてはあったほうが良いという意見もあった。今回地面ばかりの画像が抽出されることを想定していなかったが、今後はこのようなふとした一面も抽出できるような工夫も考えていきたい。

6.3.4 質問4

平均値は2.6と低くなっている。10名中7名は3未満の評価をつけており、全員が“システムの抽出枚数が少ない”という意見だった。今回の動作実験では、すべての画像を一度にシステムに通したので大まかな分類になってしまいクラス数が少なくなってしまうことが原因だと考える。今後は時間ごとに区切るなどして抽出枚数が増えるような工夫をしようと考えている。

6.3.5 質問5

平均値は3.9と高くなっている。これは主に今後への期待が大きく、自由記述ではどんな画像が抽出されるか気になるという意見や、精度次第では使ってみたいという意見があった。現時点のシステムに対しての意見は、無駄な画像が省かれるのは嬉しいという意見や、枚数が減るのは楽といった意見があった。また、自分で抽出するのも楽しいが、何千、何百という画像から抽出するのは大変という意見もあった。このことより、多めに画像を抽出し、減った候補から手動で抽出してもらうという方法も有効なのではないかと考える。

6.3.6 自由記述・その他意見

無駄な画像が省けているのは評価できるが、良い画像まで排除されてしまっているという意見が多く見られた。今後は抽出枚数を増やし、必要な画像を残すということを意識してシステムを改良していきたい。

7. まとめ

本論文では画像の自動撮影による問題点を解決するために、自動で画像を選別するシステムを提案し、試作した。大西らの研究の考察より、本研究ではバイクの運転中の記録を想定した。画像選別システムは主に3段階に分かれており、ブレ・ボケ画像の排除、類似画像の分類、類似画像からの有用な画像の抽出を行っている。ブレボケ画像の排除ではエッジ検出を行い、その分散が小さいものを排除した。類似画像の分類にはDCNNで抽出した特徴をX-meansを用いて分類した。有用な画像の選択ではSegNetを用いて空の大きな画像を抽出した。動作実験では基本的に空の大きな画像がとれており、予想通りの結果となったが、抽出枚数やクラスタリングの方法など課題が残る結果となった。評価実験では画像による振り返りが有効であることが確認できた。また、システムで抽出した画像が思い出の振り返りに有効であることも示した。しかし、全体としては無駄な画像は省けているが、よい画像まで排除されてしまっているという意見が多くみられた。今後は抽出基準の再検討が必要であると考えている。

現在の抽出基準案としては、走行中の顔の向きから興味度を算出し、興味度の高いシーンを抽出することを考えている。また顔の向き以外にも脈拍や体温など様々な生体信号を用いた興味度の算出を考えている。

参考文献

- [1] Narrative, “NarrativeCrip”, 入手先 (<http://getnarrative.com/>), (参照 2017-04-25).
- [2] GoPro.inc, “GoProHero5”, 入手先 (<https://jp.gopro.com/>), (参照 2017-04-25).
- [3] 大西杏菜 他, “振り返り支援における効率的な映像要約のための自動収集ライフログ活用法”, 情報処理学会研究報告. SPT-13(4), pp.1-6, 2015.
- [4] 笠井昭範 他, “FLAG: 位置情報を基軸としたライフログ集約システム”, 情報処理学会技術報告, vol. 2014-UBI-43, no6, pp1-6, 2014.
- [5] 杉山岳弘 他, “知識映像コンテンツ制作システムを基軸とするバイク快走支援”, 第20回人工知能学会全国大会論文集, 2006.
- [6] 姫野由香 他, “イメージスケッチを用いた観光地における印象的な景観場の特性分析”, 第38回学術研究論文発表会, 2003.
- [7] SegNet, 入手先 (<http://mi.eng.cam.ac.uk/projects/segnet/>), (参照 2017-04-25)