

思い出の楽しい振り返りのための 身体装着型カメラによる体験自動記録

木下 恵理子^{†1,a)} 小坂 真美^{†1} 藤波 香織^{†2,b)}

概要: 近年、ウェアラブルカメラを用いた自動的あるいは受動的な記録が新たなライフログの形として関心を集めている。しかし、膨大な動画・写真のデータの扱いが問題となっており、インデクシングや要約の精度評価を行う研究がなされている一方で、「思い出を記録し振り返ること」におけるユーザの満足度には言及されていない。また、多くのセンサを用いることは自然な日常風景の記録を妨げる可能性がある。よって、ユーザが楽しさを感じる場面を少数のセンサで抽出することは意義があると考えた。そこで、一人称視点のウェアラブルカメラで撮影した動画データと音声データのみを用いて、楽しい振り返りが可能な場面の検出および抽出を行う。具体的には、撮影者や場を共有した人が楽しさを感じる場面を動画から検出し、静止画として出力するシステムを開発する。本論文では、「会話風景」「盛り上がり」「興味」という3つを「楽しさ場面」として定義し、「会話風景」を除く「盛り上がり」「興味」の検出を行い、静止画切り出しを行うプロトタイプシステムを開発し、楽しさ場面判定により得られた静止画のユーザに与える感情的作用および妥当性を評価した。

キーワード: ライフログ, 体験記録, ウェアラブルシステム, 音声・動画解析

1. はじめに

近年は、GoPro[1]や Narrative clip[2], SenseCam[3]をはじめとしたウェアラブルカメラによるライフログが普及しており、従来のような能動的・自主的な撮影だけでなく、受動的・自動的な撮影が可能となっている。このような受動的なライフログでは、膨大な動画や写真のデータを扱う方法が問題となっており、現状ではクラウド上でのライフログ解析[4]や多数のセンサを用いた情報集約[5]などの研究がなされている。これらはライフログを閲覧する際の利便性や閲覧機能等の有用性向上が目的であるが、「思い出を記録し振り返ること」におけるユーザの満足度には言及されていない。膨大なデータを集約するだけではユーザの目的や意図に沿っているとはいえず、振り返りの際にユーザが楽しさを感じる事ができる場面を検出することで、はじめてユーザにとって興味深い記録が可能となると考えられる。また、多数のセンサを用いることが前提となるシステムの場合、準備や装着に手間がかかり、自然な日常風景の記録を妨げてしまう可能性がある。これらの問題を解決するためには、なるべく少ないセンサを用いた検出を行う必要がある。

本論文では、思い出を楽しく振り返るための受動的ライフログにおける、少数のセンサを用いた適切な場面(以降、「楽しさ場面」)の検出の実現を目的とする。動画には映像と音声が含まれており、静止画よりも多くの情報が得られるため、ウェアラブルカメラを用いて撮影した映像と音声のみを用いて楽しさ場面の検出を行う。これらを実現する

プロトタイプシステムを開発し、システムの出力に対する評価を行う。

本稿では、第2章で関連研究を示し本研究の位置づけを明らかにする。その後、第3章でシステムの要件定義とシステム設計を行い、その内容をもとに第4章でプロトタイプシステムの実装について述べる。第5章からはプロトタイプシステムを用いて実施した評価実験とその結果を示し、これまでの内容を踏まえ第6章で考察を述べる。最後に第7章で結論を示す。

2. 関連研究

ライフログの分野において、ウェアラブルカメラに注目した研究が進められている。Sallen, et al.は、ライフログの記憶支援の有効性に関する検証として、SenseCamを用いた記憶の想起に関する実験を行い、カメラで自発的に撮影した写真よりもウェアラブルカメラで自動的に撮影した写真のほうが過去の記憶を想起しやすいことを示した[6]。よって、本研究においても自動で撮影を行うことの有用性が示唆される。また、思い出工学とよばれる分野で、山下らは電子アルバムを見せ合うことが人とのコミュニケーションを活発にすると指摘している[7]。本研究の出力結果を他人と見せ合うことで、コミュニケーションの促進が可能であると考えられる。さらに、田高らは高齢者の思い出に意図的に働きかけ心理的安定や記憶力改善をはかる療法である回想法を取り入れたケアプログラムの効果を示した[8]。本研究で得られる静止画を蓄積し振り返ることで、高齢者の心理的安定や記憶力改善が可能であると考えられる。

また、福本らは、ライフログを自動でタグ付けするための笑顔認識デバイスを提案している[9]。眼鏡に装着したフォトフレクタで着用者の顔の筋肉の変化を検知し、「微笑み」と「笑顔」を認識するシステムである。実利用による評価の中で、相手の顔を見て会話しているときに笑顔にな

1 東京農工大学大学院 工学府 情報工学専攻
Department of Computer and Information Sciences,
Tokyo University of Agriculture and Technology
2 東京農工大学大学院 工学研究院 先端情報科学部門
Department of Computer and Information Sciences,
Tokyo University of Agriculture and Technology
a) asa.kinoko0216@gmail.com
b) fujinami@cc.tuat.ac.jp

ることが示されており、笑顔の検出により会話状態の検出が可能であることが示された。しかし、このシステムで検出可能なのは着用者の状態のみであり、カメラ撮影における被写体となる周囲の人物の状態は考慮されていない。また、日常を記録するライフログにおいては「微笑み」と「笑顔」だけでは状況として不十分であり、検出する項目を増やすことで多くの場面を検出する必要がある。本研究では、被写体の状況を考慮した判定と笑顔以外の重要場面の判定も行う。堀らは、カメラやマイクロフォン、加速度センサ、ジャイロセンサ、GPS、脳波計等を用いたライフログビデオの検索効率化システムを提案している[5]。時刻や位置、脳波の情報だけでなく、体の動きや天気、閲覧 Web サイトから収集した情報をキーとしたビデオ検索が可能であるが、センサの種類が多く、現状では設置・運用にかかる人的コストが大きいため、ユーザの行動が制限される。このため、本研究では少数のセンサを用いたライフログを実現する。

これらの研究では、検索や検出における精度を評価するものが多く、システムの出力がユーザに与える感情的な効果には言及がないものが多い。そのため、本研究では実際のライフログとしての有用性についても検証を行う。富士フィルム株式会社のアルバム作成サービスである Year Album では、写真に写っている人数や顔の向き、明るさやピントのブレなどを考慮した写真選択を行う Image Organizer 技術が用いられている[10]。しかし、この技術で選択される写真は「人がきれいに写っている写真」であり、視覚的な情報のみを用いた判定となる。そのため、印象深い場面を検出し写真を選択することはできない。Ratsamee, et al.は、画像列のライフログにおける感情的基準にもとづいたキーフレーム選択のフレームワークを提案している[10]。ここでは、皮膚電気活動 (EDA) からカメラ着用者の心理状態を推定し、キーフレーム選択の基準としている。しかし、キーフレームはシステムにより分割された場面ごとに1枚ずつのみ選択されるため、イベントとしての重要度が考慮されていない。本研究では場面分割を行わず、ごく短い区間ごとに判定を行う。

3. 楽しさ場面判定システム

3.1 システム要件

楽しさ場面を自動判定するための要件として以下の2つを設定する。

- ユーザが楽しさを感じる場面を動画から検出すること
- 手軽に思い出を振り返ることが可能な出力であること

本研究では、センサの数や種類を減少させ簡単に利用するため音声を含む動画ファイルのみを用いて場面検出を行う。このため、撮影者やその周囲にいた人が思い出として振り

返った際に楽しいと感じる場面を定義し、動画ファイルにおけるその特徴や検出手法を検討する。撮影される状況は、自宅での家族との談話や外出中の興味深い出来事などさまざまであるため、移動中でも撮影者視点での撮影が可能なウェアラブルカメラとし、あらかじめウェアラブルカメラによって撮影された動画に対して場面検出を行う。また、Jones, et al.は、動画のようなリアルタイムコンテンツを視聴しながらの集団コミュニケーションが静止画より少なく単純になる傾向があることを指摘している[12]。このことから、閲覧者のコミュニケーションを阻害せず、かつ手軽な思い出の振り返りが可能な静止画を出力とする。本システムで得られた静止画を複数人で閲覧しながら会話を行う際に、閲覧時の会話が盛り上がり、より楽しい思い出の振り返りが可能となると考えられる。

本システムは、あらかじめウェアラブルカメラによって撮影された動画に対して場面検出を行う。システム利用の概念図を図1に示す。事前に撮影した動画データをシステムに与えると、自動で判定された楽しさ場面が静止画群として出力され、閲覧や共有が可能となる。

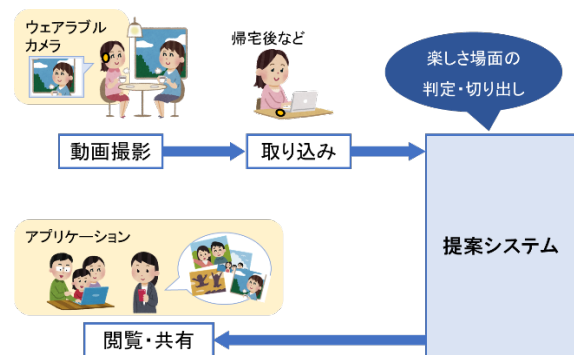


図1 システム利用図

3.2 設計

楽しさを感じる場面を定義するため、オンライン上で男性25人、女性24人の計49人にアンケート調査を実施した。回答者は94%が20~29歳である。「(A)自動で(いつの間にか)撮影してほしいと思うもの」、「(B)自動撮影では意味がないと思うもの」について以下10項目から複数回答してもらい、図2に示す結果を得た。

- (1) 旅行先で見つけた有名なもの
- (2) きれいな風景 (自然, 建物など)
- (3) スポーツを楽しんでいる様子
- (4) カフェやレストランでの食事風景
- (5) 新しい・珍しいもの
- (6) 嬉しいことがあった時の様子
- (7) 特別な行事での様子
- (8) 誰かと何気ない会話をしている様子
- (9) 一緒にいる人が楽しそうにしている様子
- (10) 全員で盛り上がった状態

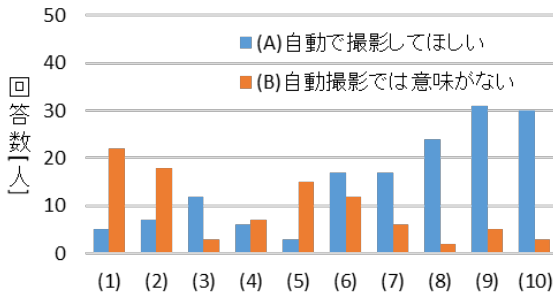


図 2 アンケート結果

図 2 から、「(1)旅行先で見つけた有名なもの」、「(2)きれいな風景」、「(5)新しい・珍しいもの」に関して自動撮影に必要性を感じない人が多く、本研究で扱うには不適切であると考えられる。一方で、「(8)誰かと何気ない会話をしている様子」、「(9)一緒にいる人が楽しそうにしている様子」、「(10)全員で盛り上がった状態」に関しては自動撮影に肯定的な回答が多く、本研究で扱う項目として適切であると考えた。よって、楽しさ場面として「会話風景」と「盛り上がり」の2つを挙げる。また、自動撮影であることを利用し、ユーザが無意識に注目したものを検出する目的で「興味」を新たな楽しさ場面として追加した。よって、本研究では以下3場面を楽しさ場面として定義する。

- (ア) 会話風景 : 会話中で特に思い出深いと感じる場面
- (イ) 盛り上がり : 笑いや大声が上がっている場面
- (ウ) 興味 : 撮影者が無意識的に注目した場面

これらの場면을判定するため、表 1 に挙げる項目を特徴として検出し、機械学習で構築した判定器を用いた場面判定を行う。

表 1 楽しさ場面と特徴

楽しさ場面	特徴	説明
(ア)会話風景	笑顔	会話している相手が笑顔になったとき
	注視	会話している対象をしっかりと見ているとき
(イ)盛り上がり	笑い声	笑い声が上がっているとき
	音量	場の音量が大きいとき
(ウ)興味	注視	撮影者が単一の物体や人物をしっかりと見ているとき

特徴量は動画像 1 秒分、60 フレームの区間で算出し、場面判定も 1 秒ごとに行われる。分類器は 3 つの楽しさ場面それぞれに対し作成し、YES (「会話風景」、「盛り上がり」、「興味」と NO (「それ以外」) の 2 クラスで分類を行う。各分類器で YES クラスである確からしさ (出力予測) を算出し、判定結果を合算した「撮影度」を静止画切り出しの

基準とする。これらを踏まえ、処理フローは図 3 のようになる。

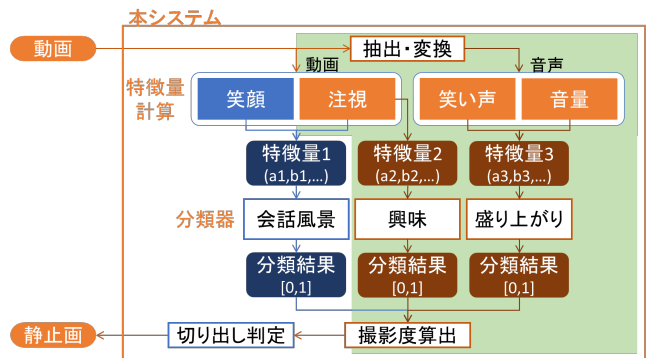


図 3 処理フロー

「笑い声」と「音量」に関しては動画像ではなく音声データを用いた計算を行うため、動画ファイルからの音声抽出を行う必要がある。また、切り出し判定は閾値による判定や上位から指定枚数を選択する方法などを想定している。

4. プロトタイプシステム実装

4.1 開発環境

本稿執筆時点では笑顔検出機能が未実装であり、「会話風景」を除いた「盛り上がり」と「興味」の2場面の判定についての実装を行った (図 3 緑背景部)。

本研究では、一人称視点での撮影が可能である Panasonic 社の A1H[13] をカメラとして使用した。また、動画や静止画の入出力等に OpenCV ver. 2.4.9[14]、音声処理の一部でフーリエ変換ライブラリ FFTW ver. 3.3.4 を用いた。A1H では MP4 形式で動画ファイルが保存されるため、音声データに関する計算を行う際に CSV ファイルへの変換を行う。本研究では、フリーソフト EcoDecoTooL ver. 1.14[15] を用いて WAV 形式の音声ファイルを抽出し、wavcsvwav ver. 2.0[16] を用いて WAV 形式から CSV 形式へ変換を行った。なお、オーディオサンプルレート 8000Hz、オーディオチャンネル数 1 での音声抽出を行った。

4.2 特徴量の検討と計算方法

4.2.1 注視

フレームごとのヒストグラムの変化から画面の動きを検出できると考え、ヒストグラムの変化を数値化し、類似度とすることで注視の度合いを数値で算出する。そのため、前処理として減色処理を行う。今回はグレースケールでの減色を行った。その後、ヒストグラムを計算し、各階調ごとの差の絶対値の総和をとり、式(a)のように注視を計算する。注視が sum 、ヒストグラムの k 階調目の値が $Hist[k]$ 、階調数が n である。

$$\text{sum} = \sum_{k=1}^n |Hist_{prev}[k] - Hist_{curr}[k]| \quad (a)$$

処理の流れを図 4 に示す。階調数やフレームの取得間隔を設定したのち、動画ファイルから読み込んだ各フレームについてヒストグラムを計算し、差を注視として保存する。

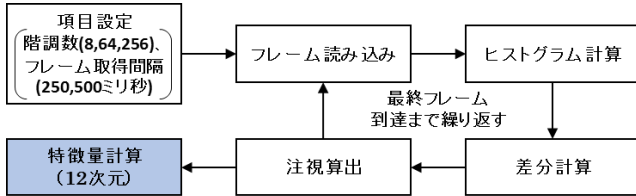


図 4 注視検出の流れ

階調数やフレームの取得間隔を変化させた計 12 次元の特徴量を用いる。注視検出に用いる特徴量を表 2 に示す。

表 2 注視の特徴量

特徴量名	説明
gray_nX_jY	X 階調, Y ミリ秒間隔での注視 (X : 8, 64, 256, Y : 250, 500)
gray_nX_jY_sub	gray_nX_jY の 1 秒前との差

4.2.2 笑い声

基本周波数の考え方をういて笑い声を検出する。音声における基本周波数は、音声に含まれる周期的成分のうち最も小さい周波数のことであり、人の声の高さを得ることが可能である。笑い声は相対的に高い音であることから、音声データから基本周波数を算出することで、笑い声を検出することが可能であると考えた。基本周波数の推定は自己相関関数を用いた。以下で述べる 2 種類のノイズ推定方法を考案し、環境音を除去した後、人の声の範囲であると考えられる 100Hz~1000Hz の範囲で自己相関関数を計算し、基本周波数を求める。ノイズの推定方法の 1 つ目は、frequency subtraction 法[17]の考え方をういた方法(「ノイズ除去」とする)である。時間的に近接している区間の周波数成分を比較し、突出している成分を非ノイズとして動的な推定を行う。2 つ目は、振幅の大きさに着目し環境音と思われる小振幅区間の振幅を縮小する方法(「スケーリング」とする)である。設定した閾値以下の値をとるデータに 0 以上 1 以下の値を比例的に乗算することで縮小を行う。笑い声検出処理の流れを図 5 に示す。前処理として環境音除去を行った、もしくは行っていない音声ファイルを入力とし、計算を行うウィンドウサイズやスライド幅を設定し、ウィンドウごとに自己相関関数から基本周波数を計算する。その後、声が高くなった区間を笑い声とするため、基本周波数の 1 秒間の上昇量から特徴量を計算する。

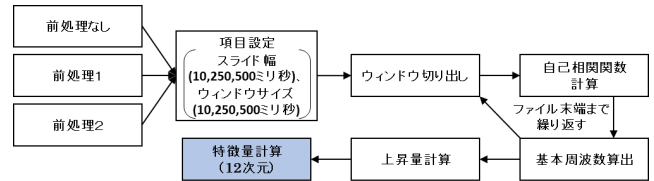


図 5 笑い声検出の流れ

計算を行うウィンドウサイズやスライド幅、ノイズ除去方法を変化させた計 12 次元の特徴量を用いる。笑い声検出に用いる特徴量を表 3 に示す。

表 3 笑い声検出の特徴量

特徴量名	説明
ff_wXsY_nomal	ウィンドウサイズ X 秒, スライド幅 Y ミリ秒 (X : 10, 250, 500, Y : 10, 250, 500)
ff_wXsY_min	ノイズ除去処理後の ff_wXsY
ff_wXsY_scale	スケーリング処理後の ff_wXsY

4.2.3 音量

波形の変位のグラフにおいて、ある短い区間ごとの絶対値の最大値をその区間の音量とした。その後、移動平均をとり平滑化を行う。音量検出処理の流れを図 6 に示す。計算を行う区間を設定し、区間ごとに音量を算出し、移動平均を計算する。区間や移動平均のデータ数を変化させた計 12 次元の特徴量を用いる。音量検出に用いる特徴量を表 4 に示す。

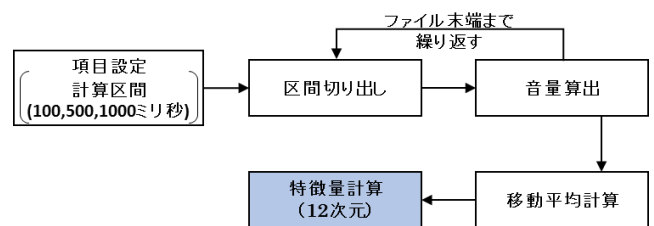


図 6 音量検出の流れ

表 4 音量検出の特徴量

特徴量名	説明
maxX_aveY	計算区間 X ミリ秒, Y データでの移動平均 (X : 100, 500, 1000, Y : 5, 7)
maxX_aveY_gap	maxX_aveY の 1 秒前との差

4.3 特徴量選択と分類器構築

前述の計 36 の特徴量を用いて、機械学習ツール Weka[18] による分類器の作成と評価を行う。12 次元を興味、24 次元を盛り上がりの分類器作成に用いる。学習データにはあらかじめ 2~6 人で撮影とラベル付けを行った 5 イベント計

160 分の動画ファイルを用いた。動画の内容と時間を表 5 に示す。

表 5 学習データの内訳

内容	時間[分]
ダーツで遊んでいる様子	50
遊園地で過ごしている様子	10
テーブルゲームで遊んでいる様子	40
カラオケで歌っている様子	30
居酒屋で飲んでいる様子	30

分類に用いるアルゴリズムは決定木数 100 の RandomForest とし、10 分割交差検定の結果、F 値が「興味」で 0.731、「盛り上がり」で 0.795 となった。また、さらなる精度向上のため、各分類器で用いる特徴量に対して属性選択を行い、特徴量を絞り込んだ。絞り込みは、Weka に掲載されている相関に基づく特徴量補集合評価法 (CFS: Correlation-based Feature Subset Evaluation) [19] と前方貪欲探索 (Forward Greedy Stepwise Search) を用いて、特徴量候補の中からクラスの識別に貢献しつつも互いの相関が弱い組み合わせを見つけた。その後、貢献度上位の属性から順番に組み合わせて検証を行った際に F 値が最も高くなる組み合わせを選択した結果、「興味」においては上位 7 個、「盛り上がり」においては上位 11 個の特徴量を用いることで、F 値がそれぞれ最高値 0.731、0.810 となった。この結果、特徴量数を半分に減らしても同等以上の分類性能を得ることができた。選択された特徴量を有効順に表 6 に示す。この表より、注視判定においては減色後の階調数が少ないほど貢献度が高く、盛り上がり判定器において笑い声に関する特徴の貢献度が著しく低く選択されていないことが分かる。

表 6 属性選択後の特徴量 (有効順)

興味	盛り上がり
gray_n8_j500	max500_ave7
gray_n8_j250	max1000_ave7
gray_n64_j250	max100_ave7
gray_n256_j250	max500_ave5
gray_n64_j500	max1000_ave5
gray_n256_j500	max100_ave5
gray_n8_j250_gap	max100_ave7_gap
	max100_ave5_gap
	max500_ave5_gap
	max500_ave7_gap
	max1000_ave5_gap

4.4 撮影度の算出

撮影度は、Weka の出力予測で得られる「興味」と「盛り上がり」の確からしさをを用いて算出する。今回は、各分類結果における確からしさを単純平均 (0.0~1.0) とする。実用においては、閾値やそれに類する判断基準を設け、撮

影するか否かを決定する必要があるが、プロトタイプシステムにおいてはこれらを定めず、撮影度とユーザの評価を比較・考察する。

5. ユーザ評価実験

5.1 実験方法

楽しさ場面判定により得られた静止画がユーザに与える感情的作用および妥当性を評価する実験を行った。実験は 3 人ずつ 3 組の計 9 人の被験者で行った。3 人は普段から会話を行う友人関係である。グループで撮影した動画に対し、システムが切り出した静止画へのユーザの印象を調査するため、インタビューを行った。さらに、ユーザ自身が動画の切り出しを行った結果とシステムの出力を比較し、システムの判定精度の評価を行った。実験は図 7 に示すように「撮影」、「インタビュー」、「ラベル付け」の 3 つのフェーズに分かれている。

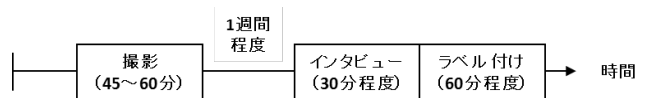


図 7 実験の流れ

● 撮影

表 7 に示す着用者の行動の性質が異なると考えられる 3 イベントに参加する様子を 15~20 分ずつ撮影する。その後、算出された撮影度をもとにシステムが静止画を切り出す。

表 7 撮影イベントとその性質

イベント	性質
散歩	<ul style="list-style-type: none"> ・会話の相手を見ていない ・視点移動が多く不規則である ・屋外での撮影である
机を囲んでの談笑	<ul style="list-style-type: none"> ・会話の相手を見ている ・視点移動が少なく不規則である ・屋内での撮影である
カードゲーム	<ul style="list-style-type: none"> ・会話の相手を見ていない ・視点移動が少なく周期性がある ・屋内での撮影である

● インタビュー

静止画で当時のことを思い出しながらインタビューを行うため、1 週間程度の期間を開けた後、切り出した静止画を被験者に見てもらい、印象を尋ねる。静止画は各イベント 30 枚ずつの計 90 枚とし、さまざまな撮影度の静止画を時系列がバラバラになるように並べ替えて提示する。

被験者は各静止画に対し自分の電子アルバムに残したいと考える度合いをそれぞれ 5 段階 (5 が最高) で評価し、理由を回答する。

● ラベル付け

被験者にカメラを装着していたイベントの動画を閲覧し

てもらい、楽しき場面に対して1秒単位でのラベル付けをしてもらう。ラベルは「盛り上がり」と「興味」であり、ラベルが付かなかった区間は「その他」とした。

5.2 実験結果

インタビュー以降の実験において1名に不備があったため、当該被験者を除いた8名分の結果を示す。

● インタビューによる印象調査と判定精度評価

システムが切り出した静止画へのユーザの印象を調査する目的で行ったインタビュー結果を示す。撮影した動画から得られた静止画群に対し被験者が行った5段階評価（以降、「評価点」）について、システムが算出した撮影度との比較を行う。なお、撮影度が0.0~1.0の値であることから、評価点も同様に0.0~1.0に正規化した。静止画の一例を図8に示す。静止画の下に撮影度と評価点をカッコ内に並べて示す。aは撮影度と評価点がともに高く、「笑顔で楽しそう」という肯定的な意見が得られた。一方、bは「笑顔で楽しそう」という意見のほかに「アングルが上にずれている」という否定的な意見が多く、評価点が下がっている。cは「人が写っていない」「特に思い出せることがない」という否定的な意見のみが得られ、撮影度評価点ともに低い値となっている。dは、撮影度は低いが「白熱してはいないが面白い写真」という意見が得られ高い評価点となった。

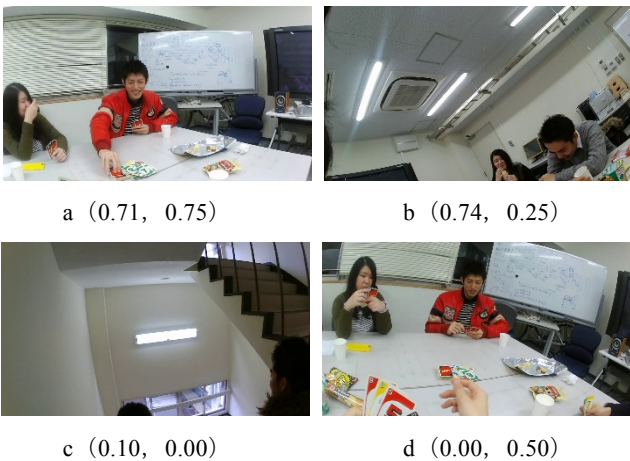


図8 静止画の一例（撮影度，評価点）

また、撮影度と回答者の平均評価点の相関係数を表8に示す。相関係数の算出に用いたデータ数は実験の中で被験者に提示した静止画の枚数と同じ30である。3組のうちグループAの「散歩」が最大値0.313、グループCの「談笑」が最小値-0.216となった。イベントごとの平均でも、「散歩」が最も相関が高く「談笑」が最も相関が低い結果となったが、全体的に相関は低かった。

表8 撮影度と評価点の相関係数

	散歩	談笑	カード
グループA	0.313	0.001	0.084
グループB	0.019	-0.050	0.039
グループC	-0.016	-0.216	0.107
平均	0.105	-0.088	0.077

また、表8に示した撮影度と評価点の相関係数において、「興味」と「盛り上がり」の各判定結果との相関係数を表9aおよび9bに示す。この結果から、散歩においては興味が、談笑やカードについては盛り上がり問わずに正の相関を示していることが読み取れる。

表9a 興味の判定結果と評価点の相関係数

興味	散歩	談笑	カード
グループA	0.391	-0.145	-0.120
グループB	0.071	-0.040	0.010
グループC	0.020	-0.261	-0.181
平均	0.161	-0.149	-0.097

表9b 盛り上がりの判定結果と評価点の相関係数

盛り上がり	散歩	談笑	カード
グループA	0.019	0.230	0.168
グループB	-0.039	-0.021	0.025
グループC	-0.048	-0.150	0.308
平均	-0.023	0.197	0.167

次に、各グループにおける全被験者の平均評価点を小数点第一位で四捨五入した1~5の5つの評価点に着目する。イベント別回答数分布は図9のようになった。このことから、「散歩」よりも「談笑」に高い点が付きやすく、「カードゲーム」はさらに高い点が付きやすいことが分かる。

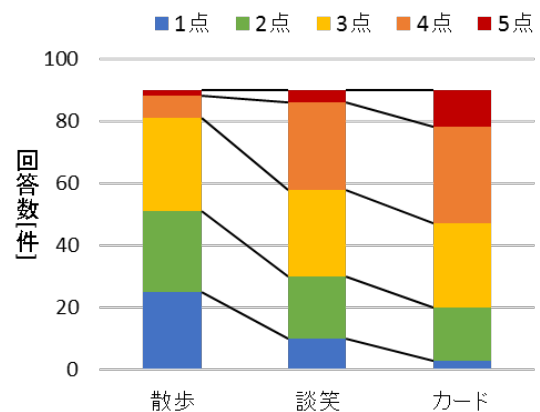


図9 イベント別評価点回答数分布

また、全員が最高点を付けた静止画に関して得られた意見として「面白い瞬間が撮影されている」、「楽しそうな様子が伝わる」、「何をしているのかよく分かる」、「(ゲーム中

の) 必死な感じが伝わる」などが挙げられる。逆に、全員が最低点を付けた静止画に関して得られた意見は「誰も写っていない」、「写真がブレている」、「アングルや構図が悪い」、「特に印象に残らない」のようになった。

● ラベル付けによる判定精度評価

ユーザ自身による動画のラベル付けの結果とシステムの出力を比較し、システムの楽しさ場面の判定精度を評価する。すべての動画に対する「興味」「盛り上がり」のラベルを集計し、ラベルの有無と撮影度との関連を示す。図 10 は、最小値、第一四分位、中央値、第三四分位、最大値を箱ひげ図で表したものである。「ラベル無」は被験者が一人も「興味」または「盛り上がり」のラベルを付けなかった区間の撮影度の分布を表しており、「ラベル有」は被験者が一人でも「興味」や「盛り上がり」のラベルを付けた区間の撮影度の分布を表している。

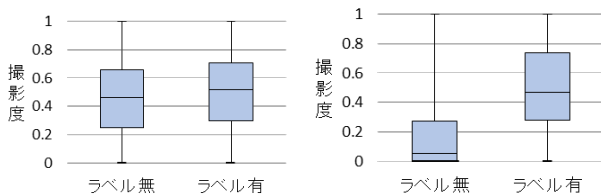


図 10 ラベルの有無での撮影度の分布
(左：興味，右：盛り上がり)

「興味」と「盛り上がり」それぞれにおいてラベルの有無に対し有意水準 5%で t 検定を行った結果、「興味」では $t(7944) = 3.72, p = 2.01 \times 10^{-4}$ 、「盛り上がり」では $t(391) = 23.46, p = 1.26 \times 10^{-76}$ となった。よって、どちらもラベルの有無による分布に有意差が見られた。とくに、「興味」よりも「盛り上がり」に関して分布に大きな差があり、盛り上がりに対するシステムの場面判定のほうがユーザの判断基準と近いということが読み取れる。

6. 考察

6.1 楽しさ場面検出の精度

表 8 で、すべての動画・イベントにおいて撮影度と評価点の相関が小さいことから、撮影度と評価点にはほぼ関連性がないと考えられる。つまり、必ずしもシステムが出力する撮影度がユーザの主観的印象に則しているとはいえない。また図 10 からは、システムの分類とユーザのラベル付けに関して、「興味」と比較して「盛り上がり」がより適切な切り出しが行われているといえるが、ラベル有りでの撮影度の中央値が 0.5 を下回っていることから、精度に関してはまだ改善の余地がある。「興味」に関しては、ユーザが前を向いた状態での直進移動や、何気なく立ち止まっている状態などが注視と判断され、「興味」として検出されて

いたと考えられる。このような「興味を伴わない注視」を取り除くための特徴量を新たに考案・検討する必要がある。このような判定精度の低さも原因の一つであるが、被験者からの意見で判明した次の項目も考慮する必要がある。

- ・ 場面の珍しさ
- ・ アングルや構図
- ・ 画面のブレ
- ・ 写っている人数
- ・ 他の出力結果との類似度

ここでの珍しさは、「普段ふざけない人が面白い発言をした」、「ゲームが不得意な人が勝利した」などの時間的なものであり、「新装開店した飲食店での飲食」や「面白い形の建造物がある」のような空間的な珍しさとは区別される。また、撮影度と静止画の評価点の関係について、図 8a や c のように撮影度と評価点が近い結果となる静止画が得られる一方で、b や d のように撮影度と評価点が大きく異なる結果となる静止画も得られている。b は盛り上がっているがアングルが悪いためにユーザの印象が悪くなり、d は盛り上がりはしていないが人数が多いためユーザの印象が良くなっている。さらに、類似する静止画が複数枚提示された際に、2 枚目以降の評価点が低下する傾向があったため、類似度の高い静止画を除く処理が必要となる。よって、上に箇条書きで示した新たな項目を特徴として追加することで精度が改善されると考えられる。たとえば、加速度センサから取得したカメラの向きをアングルの特徴量としたり、エッジ検出からブレの生じているフレームを検出したりする方法が考えられる。また、顔検出や人体検出を用いた人数推定を行うことで写っている人数を考慮することが可能である。

6.2 イベントの楽しさと振り返り時の楽しさの関連

図 9 から、散歩、談笑、カードゲームの順で高得点が付きやすくなっていることが読み取れる。散歩は、見慣れた風景の写真や後姿の写真が多かったために評価点が低くなってしまったことが考えられる。一方で、カードゲームは散歩や談笑よりもイベント自体を楽しんでおり、笑顔の写真が多く得られ、評価点が高くなったと考えられる。また、カードゲームに関して、写りこんでいるカードから当時の状況を想起できたという意見があり、状況を思い出す手がかりの数が評価点に影響する可能性がある。

6.3 撮影度の算出方法

表 9a および 9b から、散歩のような移動の多い状況では「興味」、談笑やカードゲームのような動きの小さい状況では「盛り上がり」がわずかながらユーザの印象に則した判定をしていることが判明した。これは、動きの少ない場面

で検出される注視よりも常に動いているような場面で検出される注視のほうが重要であるためであると考えられる。また、談笑やカードゲームにおける注視度は常に大きくなりやすく、本当に興味をもって注視している区間の検出が困難となっている。よって、カメラ着用者の動きが多い状況では「興味」、動きが少ない状況では「盛り上がり」の影響が大きくなる可能性が示唆される。本研究における撮影度はすべての楽しさ場면을等しい重みで加算していたが、撮影状況に合わせて動的に楽しさ場面の重み付けを行うことでより高い精度での撮影度算出が可能となると考えられる。

7. おわりに

本研究では、思い出を楽しく振り返るための受動的ライフログにおける適切な場面の検出及び抽出を行うシステムのための、適切な場面の定義やその検出手法の検討および評価を行った。動画データと音声データを用いて「興味」「盛り上がり」の特徴を算出し、機械学習での分類によって1秒単位での判定を行い、静止画を切り出した。得られた静止画に対する印象調査やユーザによるラベル付けとの比較を行い、次の成果および知見を得た。

- 音量を用いて盛り上がり状態を有意に判定することができる
- 静止画に対する印象は、アングルや構図、ブレの有無により変化する
- 静止画に対する印象は、珍しさや状況の伝わりやすさにより向上する
- 状況により思い出として残したいと感じる項目が変化する可能性がある

今後の課題として、改善すべき点は以下である。

- 状況の伝わりやすさを考慮した判定場面の再検討
- ブレやアングル、構図を考慮した特徴量の再検討
- 状況に則した判定結果の動的重み付け手法の考案

これらの課題を解決し、思い出を楽しく振り返ることができる楽しさ場面判定システムの要件を満たしたシステムの実現を目指す。

謝辞

本研究の一部は、日本学術振興会科学研究費補助金ならびに文部科学省特別経費「持続可能社会に向けた知的・情報空間技術の創出」によるものである。

参考文献

- [1] Woodman Labs; “GoPro”, <http://jp.gopro.com/>
- [2] Narrative; “Narrative Clip 2 – The world’s most wearable camera” <http://getnarrative.com/>
- [3] Microsoft “SenseCam”
<http://research.microsoft.com/en-us/um/cambri/dge/projects/sensecam/>
- [4] 黒崎裕子, 山下暁香, 小口正人. “オンライン機械学習フレームワーク Jubatus によるライフログからの情報抽出.” DEIM2014, (2014).
- [5] 堀鉄郎, 相澤清晴. “ライフログビデオのためのコンテンツ推定 (画像符号化・通信・ストリーム技術, 及び一般).” 電子情報通信学会技術研究報告. IE, 画像工学 103.514 (2003): p.67-72.
- [6] Sellen, Abigail J., et al. “Do life-logging technologies support memory for the past?: an experimental study using sensecam.” In Proceedings of the SIGCHI conference on Human Factors in Computing Systems. ACM, (2007): p. 81-90.
- [7] 山下清美, 野島久雄. “思い出コミュニケーションのための電子ミニアルバム提案.” ヒューマンインタフェースシンポジウム 1 (2002): p261-264.
- [8] 田高悦子, 金川克子, 立浦紀代子, 和田正美. “在宅痴呆性高齢者に対する回想法を取り入れたグループケアプログラムの効果.” 老年看護学: 日本老年看護学会誌: journal of Japan Academy of Gerontological Nursing, 5(1), (2000): p.96-106.
- [9] 福本くらら, 寺田努, 塚本昌彦. “ライフログにおける自動タグ付けのための笑顔認識機構の設計と実装.” 研究報告モバイルコンピューティングとユビキタス通信 (MBL) 2013.18 (2013): p.1-8.
- [10] 胡学斌, 高森哲弥, 山路啓. “大量画像からのフォトブック作成を簡単・手軽にする Image Organizer 技術.” Imaging Conference Japan 論文集. 日本画像学会, 2014.
- [11] Ratsamee, Photchara, et al. “Keyframe Selection Framework Based on Visual and Excitement Features for Lifelog Image Sequences” International Journal of Social Robotics, Volume 7, Issue 5, pp859-874
- [12] Jones, Quentin, et al. “Information overload and the message dynamics of online interaction spaces: A theoretical model and empirical exploration.” Information systems research 15.2 (2004): p.194-210.
- [13] Panasonic “A1H”
<http://panasonic.jp/wearable/a1h/>
- [14] OpenCV <http://opencv.jp/>
- [15] EcoDecoTool
<http://osdn.jp/projects/ecodecotool/>
- [16] wavcsvwav
<http://sunfieldkikaku.web.fc2.com/newpage9.html>
- [17] 大槻典行, 宮永喜一. “音声認識に効果のある周波数領域での音声雑音除去: ランニングスペクトルフィルタ (RSF) の効果 (認識・理解・対話).” 電子情報通信学会技術研究報告. SP, 音声 103.93 (2003): p.31-35.
- [18] The University of Waikato “Weka”,
<http://www.cs.waikato.ac.nz/ml/weka>
- [19] Hall, Mark A.: Correlation-based Feature Selection for Machine Learning. PhD thesis, The University of Waikato, (1999).