

類似写真の閲覧がもたらす話題生成と対面コミュニケーション

How face-to-face communication with photo similarities effects conversations

香取真知子^{*1}

Machiko Katori

瓜生大輔^{*1*2}

Daisuke Uriu

稲見昌彦^{*1*2}

Masahiko Inami

檜山 敦^{*1*2}

Atsushi Hiyama

^{*1} 東京大学大学院情報理工学系研究科

Graduate School of Information Science and Technology, the University of Tokyo

^{*2} 東京大学先端科学技術研究センター

Research Center for Advanced Science and Technology, the University of Tokyo

This paper introduces a face-to-face communication platform using images. Our iPhone application, used by plural users simultaneously, will display a set of images with an algorithm calculating image similarities. The app accesses each user's picture library linked with Google Photos, sorts in the relevance with the algorithm, and selects which to display. This paper illustrates the app design and the validation study of the user communication. The study recruited participants divided into two groups: one that are close friends, and another are complete strangers. We intended the former group to catch up and find some new topics to talk about, while the latter group to find some things in common that might trigger communication. This pilot study reveals how the current system effects the conversations, states how this platform could articulate face-to-face communications, and finally discusses how this platform could be improved in the future.

1. はじめに

SNS 上での写真・イメージを用いたコミュニケーションが広く普及する今日 [Malik 16], SNS の利用は他者との人間関係の維持にも重要な役割を果たしている ([Chen 11] など). このような写真を介したオンラインでのコミュニケーションの需要に対し, 本研究では, オフラインでの, すなわち直接会った際に特化した写真閲覧経験に着目する. SNS 上での投稿は, フォロワーを含めた不特定多数に対しての発信なのに対し, オフラインでの会話は, その場にいる相手やグループの趣味・嗜好に特化・依存する. このような使用者の特性に基づいて, 本研究では共通性・類似性のある写真の提示方法を検討した. 類似する写真の閲覧を通し, 初対面同士では互いの興味・関心の共通点の発見を, また元々親しい者同士では円滑な近況報告を, 結果として様々な話題を誘発できるのではないかと考えた.

本研究において開発中の iPhone アプリケーション (以下, アプリとする) は図 1 のように, 複数の iPhone を並べた状態で使用する. 個々の使用者 (本図の場合は 3 人) が所有する Google Photos 上に格納された写真から, 本研究で開発中の写真提示アルゴリズムにより, 類似度の高い写真の組から順番に 20 組表示できる. また使用者が画面をスワイプすることで次の写真の組に切り替わる. 本論文では, 2018 年 10 月から 2019 年 1 月にかけて開発したアプリとアルゴリズムを用いた写真閲覧システムを準備し, 試用と検証を行なった結果と考察を報告する.

2. 関連研究

本研究は, 個人が所有している写真を利用した記憶の共有や回想を支援する研究領域 ([Uriu 09], [Nguyen 16] など) に対し, 複数の使用者が所有する類似した写真の同時閲覧に関する知見を提供する. 本アプリでは視点や構図が固定された一人称視点からのライフログ写真 [Wang 13] ではなく, よりバリエーシ

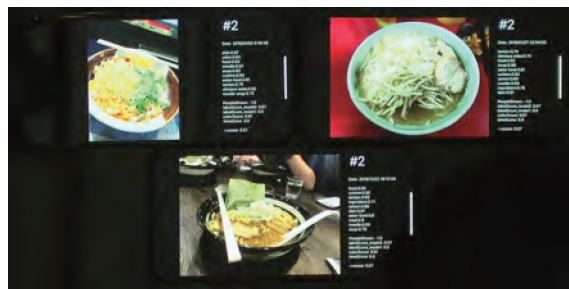


図 1: 実際の表示例. 体験者 3 名が麺類を食べた際の写真.

ン豊かなイメージの提供を目指し, 能動的に撮影された写真によって得られる体験に注目した. その際, 画像検索分野で注目されている semantic gap [Liu 07] を補完する物体認証のモデルを活用して, 雑多で大量な写真リストから共通性のある組み合わせを導く. 現在, SNS 上の投稿文章や画像を分析することでユーザの興味や趣味嗜好を特定する研究がある ([Kosinski 13] など) が, 本研究では実際に対面した際に共通の興味や趣味嗜好を発見できるイメージの提供に注目した.

3. アプリ設計

3.1 アプリの構成・基本操作

アプリは 2 人以上の体験者が対面した状態で動作する. その際, 類似・共通する写真の組を 20 組, 類似度の高い順に鑑賞できる. 写真には #1 から #20 までの番号が振られており, iPhone の画面を上下にスワイプすることで, 前後の組の写真に遷移する. 1 台の iPhone をスワイプすると Peer to Peer 通信により信号が送られ, 他の iPhone の写真も遷移する. これにより, 常に同じ番号, すなわち正しい組の写真同士が同時に表示される.

3.2 Google API の利用

本研究では, Google が提供する Photos Library API および Cloud Vision API を利用している. Photos Library API は, Google

連絡先: 香取真知子, 東京大学稲見・檜山研究室, contact@star.rcast.u-tokyo.ac.jp

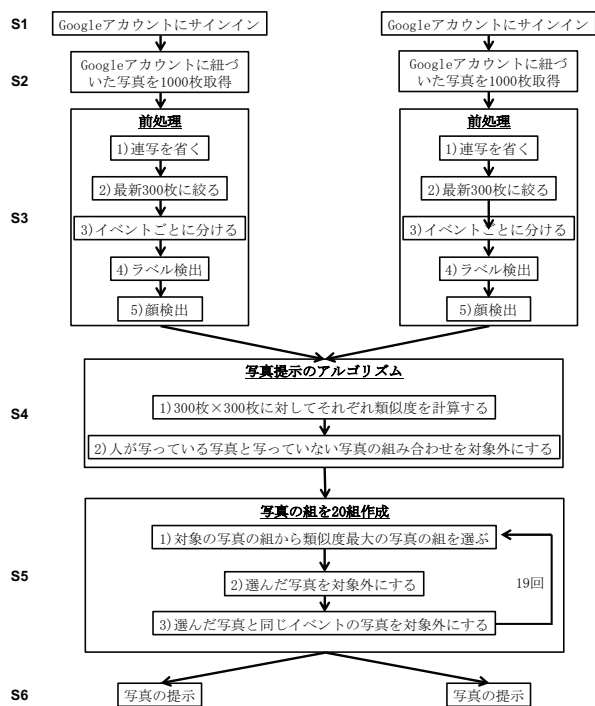


図 2: アプリの処理フロー

Photos 中の写真、動画および紐付けられたメタデータを取得・活用するための API である。Cloud Vision API は、事前にトレーニングされた機械学習モデルを活用し、画像の物体認証を行うことができる。これらを利用することで、本研究プロセスで改めて事前学習などを行うことなく画像からさまざまな情報を引き出せる。本研究では、主に「ラベル検出」と「顔検出」の API を使用した。ラベル検出は、画像の物体認証を行い、検出されたラベルの名称とその信頼度 (0~1) を取得可能な API である。顔検出は、画像に含まれる (複数の) 人物の顔を検出する API である。ただし、個人を特定する顔認証には対応していないため、本研究では被写体に人物が含まれるか否かの判定に用いている。

3.3 アプリが行う処理

アプリが行う処理の流れを図 2 に示す。まず Photos Library API を使用して、アカウント (S1) に紐づいた写真を 1000 枚取得する (S2)。次に前処理 (S3) を行う。まず、1 分以内に撮影された連続写真のうち最後に撮影された 1 枚以外を省くことで枚数を減らす (S3.1)。この処理を行ったあと、残りの枚数が 300 枚以上の場合、さらに最新 300 枚に絞る (S3.2)。さらに、全ての写真に対して次の写真が撮られるまでの時間が 30 分以上開くところで区切り、区切られた時間内に収まる写真群をグループとして記録した。本論文ではこのグループを「イベント」と呼ぶ (S3.3)。Cloud Vision API を利用してラベル検出 (S3.4) と顔検出 (S3.5) を行う。この際、取得されるラベルの最大数を 10 と設定した。本論文では、写真 I_i に対するラベル検出の結果を $L_i = \{l_{i1}, l_{i2}, \dots, l_{iU}\}$ とする (ただし $i, U \in \mathbb{N}, i \leq 300, U \leq 10$)。また、 u 番目のラベルの名前を key_{iu} 、ラベルの信頼度を $score_{iu}$ とし、 $I_i = (key_{iu}, score_{iu})$ と表す。

前処理 (S3) 後、写真提示のアルゴリズムを適用する (S4)。M 人で使用する場合、前処理によって絞られた写真の枚数をそれぞれ $N_1, N_2, \dots, N_M$ とする。このとき、計算対象となる写真の組み合わせの数 N は、

$$N = \prod_{m=1}^M N_m$$

と表せる。全ての組み合わせについて、類似度 F を計算する。 F は関数 f を用いて以下のように計算される。

$$F(I_{1n_1}, I_{2n_2}, \dots, I_{Mn_M}) = \frac{\sum_{i=1}^M \sum_{j>i}^M f(I_{in_i}, I_{jn_j})}{M C_2} \quad (1)$$

ただし、 m 番目の人が保持する写真を $I_m = \{I_{m1}, I_{m2}, \dots, I_{mN_m}\}$ と表し、 $m, n_m \in \mathbb{N}, n_m \leq N_m, m \leq M$ とする。写真 I_i と I_j における $f(I_i, I_j)$ は以下 (1.~4.) のように計算される。ただし、写真 I_i のラベル検出結果を $L_i = \{l_{i1}, l_{i2}, \dots, l_{iU_i}\}$ 、写真 I_j のラベル検出結果を $L_j = \{l_{j1}, l_{j2}, \dots, l_{jU_j}\}$ とする。

1. 写真 I_i, I_j のそれぞれに対するベクトル $l_i, l_j \in \mathbb{N}^{20}$ を作成。
2. l_i の u 成分の値を $score_{iu}$ にする ($u \in \mathbb{N}, u \leq 10$)。
3. L_j の各ラベル l_{jv} に対して、 $key_{iu} = key_{jv}$ となるようなラベルが L_i に存在し、そのラベル名が document, handwriting, line, screenshot, whiteboard, product, angle, text 以外ならば、 l_j の u 成分の値を $score_{jv}$ にする。存在しなければ、 l_j の $(10 + v)$ 成分の値を $score_{jv}$ にする。
4. 以下のように $f(I_i, I_j)$ を計算する。

$$f(I_i, I_j) = \frac{l_i \cdot l_j}{10}$$

この関数 f を用いて、類似度 F は式 (1) のように計算される。上記 3. にあるように、ノートや黒板、ホワイトボードの写真などを表す document, handwriting, line, whiteboard, angle, text といったラベルを省いて処理する。また、screenshot, product は、多くの写真に付与されるため省いた。関数 f では l_i と l_j の内積を $10 (= \text{取得されるラベルの最大数})$ で割り、値を規格化する。この処理により、一致するラベルの数が多く、かつ信頼度が高い場合に類似度が高く計算される。次に、顔検出の結果を反映させる (S4.2)。顔検出により、全 N 組の中から、人物が写っている写真同士と人物が写っていない写真同士の組み合わせのみを対象に、以下の 1.~3. の処理を繰り返し 20 組を選択し、類似度 F の高い順に #1 から #20 の番号を付与する (S5)。

1. 対象から最も類似度 F の高い組み合わせを選ぶ
2. 選ばれた組に含まれる写真を対象外にする
3. 選ばれた写真と同一イベント内の写真を対象外にする

同一イベントに含まれている写真は 1 枚のみしか提示されないようにすることで、話題性を保つことができると考えた。最後にそれぞれのアプリ画面に写真が表示される (S6)。

4. 検証

本研究では、計 14 名の協力者を集い、計 10 組 (初対面同士: 4 組, 親しい者同士: 6 組) による試用と検証を行った。前章で述べた写真選択のアルゴリズムを用いた modeA と、ランダムで表示する modeB の 2 種類で行った。体験者には、事前に Google Photos に 1000 枚以上の写真のアップロードを依頼した。ただし、どうしても他人に見せたくない写真は任意で省いてもらった。体験者には、各 mode の全 20 組を全て必ず見て会話するよう伝え、それぞれの組に対する会話の所要時間は自由とした。また、各 mode の意味は伝えなかった。

すべての検証は動画で記録し、終了後にはアンケートを依頼した。実際に提示された組の写真を振り返りながら、計 40 組の写真の組み合わせについて話題性と類似性についての 5 段階評価と、それぞれの評価に対して任意で理由などを記述してもらった。計 10 組の検証、合計 400 組の写真を表示する過程にお

いて、何らかのシステムエラーにより正しく表示されない現象が13組発生した。このようなデータについては、以下に述べる検証結果・分析・考察からは省く。

表 1: 各 mode の類似性と話題性の平均値と相関係数

		初対面グループ		親しいグループ	
平均値	類似性	modeA	modeB	modeA	modeB
	話題性	3.84	3.17	3.65	3.04
相関係数		0.31		0.13	

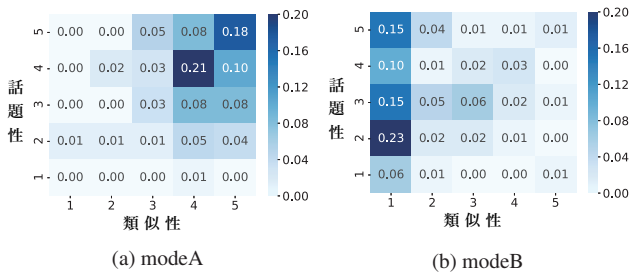


図 3: 初対面グループにおける類似性と話題性のヒートマップ

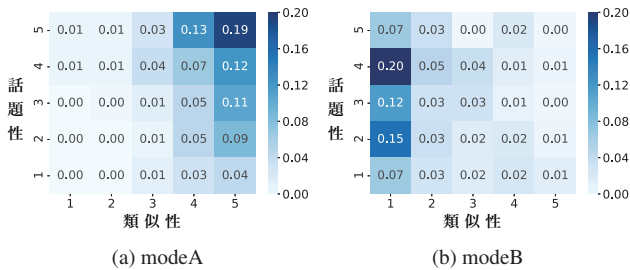


図 4: 親しいグループにおける類似性と話題性のヒートマップ

5. 検証結果・考察

5.1 初対面グループにおける類似性と話題性の相関

各グループ、各 mode で類似性と話題性の平均値と、各グループにおける類似性と話題性の相関係数を表 1 に示す。また、分布のヒートマップを図 3, 4 に示す。初対面グループにおける類似性と話題性の相関は 0.31 と、若干の相関が確認できたが、図 3 の通り、類似性評価が低いものの話題性の値が高い場合もあった。とりわけ話題性が下がってしまう 2 つのパターンがあった。1 つ目は構図や被写体がほとんど同じ写真が複数回表示される場合である。たとえばある検証では、料理の写真が提示される回数が多く、これにより途中から話題性が落ちてしまった。2 つめは、写真が撮影された背景など（以下、バックストーリーとする）が同じものである。今回の検証では、同じイベント内の写真が提示されないようにすることで（S5.3）、バックストーリーが同じ写真が複数回提示されることを防いだ。しかし、旅行など複数日にわたる出来事には多数のイベントが含まれる。体験者の多くは、これらを全て同じバックストーリーを持つ写真だと捉え、それらが複数回提示される場合に話題性評価が下がる傾向が見られた。一方、modeB においても、話題性評価が平均値の 3 を超えた（表 1）。図 3b から類似性が 1（全く類似していない）のときの話題性がバラバラであることがわかる。これは、似ていない写真の組だとしても、体験者それぞれが写真について話したい内容がある場合は話題性があったと考えられる。

5.2 親しいグループにおける類似性と話題性の相関

親しいグループでは、いずれか 1 枚でも特徴的なバックストーリーを持つ写真が含まれていれば、全員で盛り上がり、話題性評価も高くなる現象が見られた。これは親しいグループ特有のもので、初対面グループでは 2 枚の写真のいずれか一方のみに特徴的なバックストーリーがあり、もう片方にはない場合、バックストーリーを聞かされた体験者側の話題性評価が上がる一方、話した側の話題性評価は上がらない傾向があった。また親しいグループでは、本検証以前に日常的に特定の写真や互いの近況を共有していることがあった。このような既視感のある写真が含まれている場合、写真の内容を知っているもしくは聞かなくても推測できるため、話題性が落ちる現象が見られた。これらの理由から、親しいグループでは話題性が生まれる根拠が不確定的であり、結果として類似性と話題性の相関は見られなかった（表 1）。

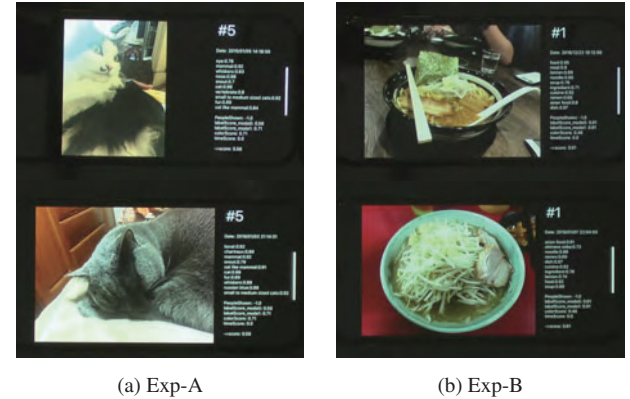
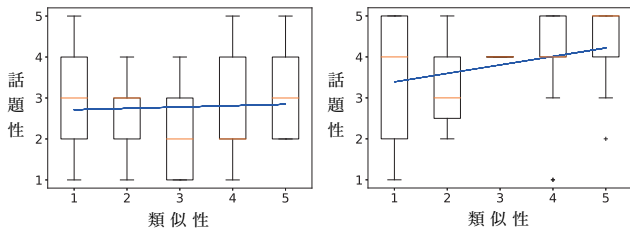


図 5: 親しいグループ (Exp-A, Exp-B) において連続して類似する写真が表示された例

5.3 類似性と話題性評価のバラツキ

次に、2 組の親しいグループでの検証において、それぞれ被写体が似た写真が連続して表示されたケース（Exp-A, Exp-B とする、図 5）について述べる。箱ひげ図上の青線は回帰直線を表しており、Exp-A では相関がない一方で Exp-B では正の相関があることがわかる。Exp-A の体験者 2 名は、ともに自宅で猫を飼っており、猫が被写体となった写真ばかりが表示された。結果として途中で猫の話に飽きてしまった。一方 Exp-B は、アルバイト先の同僚同士である体験者たちによるものである。ここでは最近食べたラーメンの写真ばかりが表示されたが、最後まで店についての情報交換を楽しんでいた。このような話題性の持続性の差は、体験者間の関係性によって引き起こされていると考えられる。Exp-A の体験者間の付き合いは Exp-B の体験者間より長く、普段から頻繁に連絡を取り合い、猫の話を含め密に情報交換を行なっている。一方、Exp-B の体験者は日常的な情報交換はそれほどなく、大量に連続するラーメンの写真が、むしろ新たな話題を誘発したと考えられる。

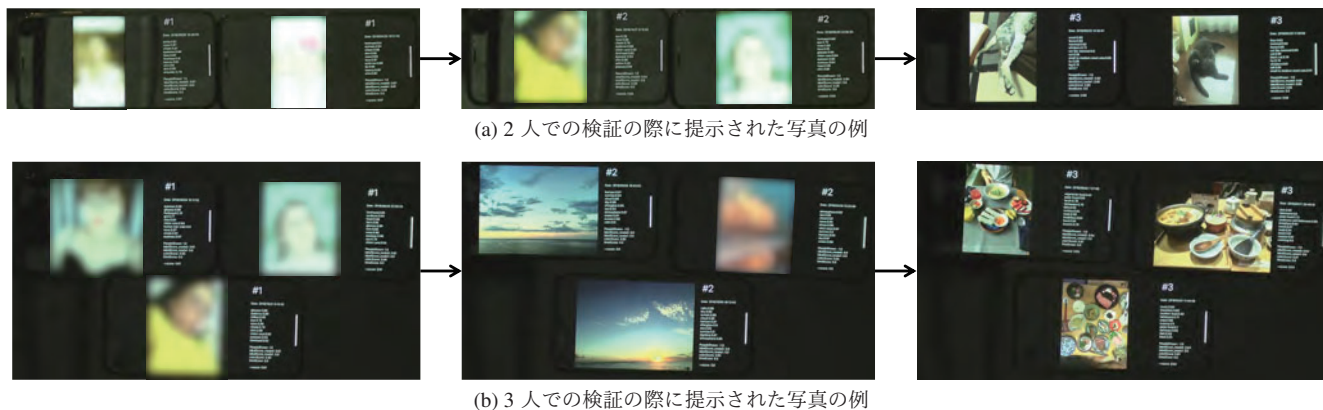


図 6: 親しい友人同士の検証 (Exp-A) において 2 人での体験から 3 人目が加わった例. 左から順に #1, #2, #3.

5.4 相手による経験の違い

体験する相手の組み合わせにより表示される写真が変化する本アプリの特徴を表す例として、相手を変えて 2 回の検証に参加した体験者の例を図 7 に示す. このように、体験を行う相手により、提示される写真は異なり、結果として誘発される話題も異なる. また、親しい 3 人グループにおける検証 (Exp-A) において、2 人で検証した場合と、もう 1 名加わり 3 人で体験した際に表示された写真 (#1~#3) の例を図 6 に示す. 2 人の類似度・共通度に基づき表示されていた図 6a に対し、3 人目が加わった図 6b では、3 人に類似・共通する風景や料理の写真に変化した. 一方で、図 6a 中央と図 6b 左のように、2 人での検証で提示された組に 3 人目の写真が加わることがある. 式 (1) からわかるように、3 枚の写真の類似度 F はそれぞれの 2 枚から計算される類似度 F の平均値であり、2 人での類似度 F が高かった組み合わせは、3 人でも出やすい. しかし、このような場合でも、使用者の関係性にに応じて誘発される会話は異なった.



図 7: 初対面グループの体験者のメンバー変更による影響の例. 上の iPhone は同一体験者, 下は別の体験者

6. 結論

本研究では、実際に対面した際に会話を促進するプラットフォームの提供を目的とし、類似性の高い写真の提示により会話の誘発をうながす、複数人が集まって使用するアプリを開発した. またアプリを使用することでどのような体験が得られるかを検証した. 検証の結果、初対面グループでは類似性と話題性に相関が見られ、類似度の高い写真によって話題が生成されることが確認できた. しかし、同じ内容の写真や 1 つの長期的な出来事に含まれる写真が複数回提示されてしまい、似ているにも関わらず話題性が下がることがあった. これを防ぐために、被写体の類似した写真の選別や、イベント (一定の時間内に撮影された複数の写真) の分け方の改良が必要である. 一方で

親しいグループでは各検証により体験者の経験が異なったが、これは親しい同士の人間関係に依存していると考えられる. 同僚、友人、家族、恋人など、親しいグループの定義をより細分化した上での検証と様々な人間関係に特化した開発が求められる.

謝辞

本研究の一部は、JST ERATO (JPMJER1701)、および、JST 未来社会創造事業 (JPMJMI18CA) の支援を受けたものである.

参考文献

- [Chen 11] Chen, G. M.: Tweet this: A uses and gratifications perspective on how active Twitter use gratifies a need to connect with others, *Computers in Human Behavior*, Vol. 27, No. 2, pp. 755 – 762 (2011)
- [Kosinski 13] Kosinski, M., Stillwell, D., and Graepel, T.: Private traits and attributes are predictable from digital records of human behavior, *Proceedings of the National Academy of Sciences*, Vol. 110, No. 15, pp. 5802–5805 (2013)
- [Liu 07] Liu, Y., Zhang, D., Lu, G., and Ma, W.-Y.: A survey of content-based image retrieval with high-level semantics, *Pattern Recognition*, Vol. 40, No. 1, pp. 262 – 282 (2007)
- [Malik 16] Malik, A., Dhir, A., and Nieminen, M.: Uses and Gratifications of digital photo sharing on Facebook, *Telematics and Informatics*, Vol. 33, No. 1, pp. 129 – 138 (2016)
- [Nguyen 16] Nguyen, V.-T., Le, K.-D., Tran, M.-T., and Fjeld, M.: NowAndThen: A Social Network-based Photo Recommendation Tool Supporting Reminiscence, in *Proceedings of the 15th International Conference on Mobile and Ubiquitous Multimedia*, MUM '16, pp. 159–168, New York, NY, USA (2016), ACM
- [Uriu 09] Uriu, D., Shiratori, N., Hashimoto, S., Ishibashi, S., and Okude, N.: CaraClock: An Interactive Photo Viewer Designed for Family Memories, in *CHI '09 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '09, pp. 3205–3210, New York, NY, USA (2009), ACM
- [Wang 13] Wang, P. and Smeaton, A. F.: Using visual lifelogs to automatically characterize everyday activities, *Information Sciences*, Vol. 230, pp. 147 – 161 (2013)