

一人称視点カメラを用いた 思い出の楽しい振り返りのための体験自動記録

木下恵理子[†] 藤波香織^{††}

東京農工大学 工学府 情報工学専攻[†] 東京農工大学 大学院 工学研究院 先端情報科学部門^{††}

1. はじめに

近年、ウェアラブル型カメラを用いた自動的または無意識的に記録を行うライフログへの関心が高まっている。それに伴い、画像や映像によるライフログにおけるセンサ情報を用いた場面推定や要約を行う研究が多くなされている [1][2]。一方、「思い出を振り返る楽しさ」に着目した研究は少なく、多くのセンサ利用は自然な記録の妨げとなる。そこで本研究では、一人称視点で撮影された映像や音の情報のみを用いて、撮影者や場を共有した人が楽しさを感じる場面(楽しさ場面)の自動判定手法を開発する。ここで、システム概要を図 1 に示す。本稿では、既報 [3][4] で定義した「会話風景」「盛り上がり」「興味」の 3 つの楽しさ場面にに対し、視覚的な評価を行う機能である「視覚評価部」と、似た出力を除外する「類似画像除外部」の実装や評価について述べる。

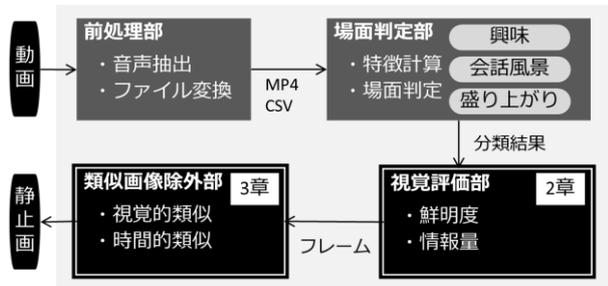


図 1 システム概要と本稿でのフォーカス

2. 視覚評価部

2.1 手法概要

既報 [3][4] より、カメラ着用者の体の動きによる画面のブレが顕著なフレームや、天井や空などを映しており情報が少ない構図のフレームの評価が下がることが課題となっていた。そのため、場面判定で得られた楽しさ場面の区間において、画面の「鮮明度」や「情報量」を評価する必要がある。視覚評価部では、これら 2 項目のスコアを計算し、視覚的に質の高いフレームを選択する。

(i) 鮮明度

鮮明度の算出には、OpenCV の Canny 関数でのエッジ検出の結果を用いる。Canny 関数からは、0 と 255 の 2 値でエッジを表した二次元配列が得られる。そこで、式(1)のようにエッジの多さを算出し、0 から 1 に正規化した値を鮮明度とした。正規化の際、縞状にエッジが現れる状態が最大と考え、分母を 2 で除算している。

$$\text{鮮明度} = \frac{255 \text{ のピクセル数}}{\text{全ピクセル数} \times \frac{1}{2}} \quad (1)$$

(ii) 情報量

情報量の算出には、画面の色のエントロピーを用いる。色のエントロピーが高いほどユーザが得る情報が増え、視覚的な印象が

向上すると考えた。エントロピーを用いて絵画の色彩評価を行った研究 [5] を参考にして、HSV 色空間のうち色相を示す H 座標と明度を示す V 座標を用いたグルーピングを行う。彩度 S が小さい画素では色相 H で 6 グループに、彩度 S が大きい画素では明度 V で 4 グループに分け、計 10 グループに振り分ける。

3. 類似画像除外部

3.1 手法概要

視覚評価部で得られたフレーム群に対し、閲覧ユーザが「別フレームと内容が酷似している」と感じるフレームを除外する。内容の酷似には、フレームの見た目そのものが似ている場合と、撮影された時刻が近い場合の 2 つが考えられる。そのため、図 2 のように視覚的類似と時間的類似を計算し、その合計から除外処理を行う。これにより、視覚的に類似した別場面の差別化や、視覚的には類似していないが思い出す内容が同一で残す必要がないフレームの除外が可能となると考えた。

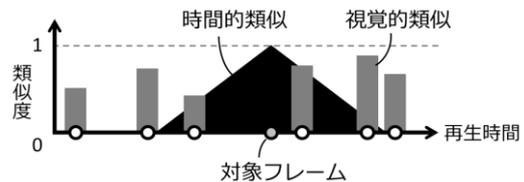


図 2 類似画像除外のイメージ図

3.2 視覚的類似

視覚的類似は、2 フレーム間でのカラーヒストグラムの差を 0 から 1 の実数値として点数化したものである。光の加減やノイズ等による色味の違いを吸収するため、前処理として画素値を定数 n で除算し、簡易的に減色を行う。Image(i, j) を i 行 j 列目の画素値、整数 N を減色数の 3 乗根とすると、減色後の各チャンネルの画素値 image'(i, j) は式(2)で表される。

$$\text{image}'(i, j) = \frac{\text{image}(i, j)}{n} \times n \quad \left(n = \frac{256}{N} \right) \quad (2)$$

減色後は、OpenCV の calcHist 関数を用いてヒストグラムを計算し、視覚的類似度の計算を行う。まず、現在のフレームのヒストグラムとひとつ前のフレームのヒストグラムの、各階調ごとの差の絶対値の和をとる。その後、全ピクセル数 $\times 2 \times$ チャンネル数で除算し、0 から 1 の値に正規化する。ヒストグラムの差は大きいほど類似していないことを指すため、最後に正規化した値を 1 から減算して視覚的類似度とする。ヒストグラムの k 階調目の値を Hist[k]、階調数を n 、全ピクセル数を p とすると、各チャンネルにおける視覚的類似度 S_v は式(3)のように表される。

$$S_v = 1 - \frac{\sum_{k=1}^n |\text{Hist}_{prev}[k] - \text{Hist}_{curr}[k]|}{2 \times p} \quad (3)$$

減色処理と同様に、RGB それぞれのチャンネルに対し計算を行い、各チャンネルの S_v の合計値をチャンネル数の 3 で除算した 0 から 1 の値を視覚的類似度とする。

3.3 時間的類似

時間的類似は、フレーム番号と動画のフレームレートから時間

Automatic Recording for Enjoyable Recall by a Wearable Camera

Eriko KINOSHITA[†], Kaori FUJINAMI^{††}

^{†,††}Department of Computer and Information Sciences, Tokyo University of Agriculture and Technology

的距離を求め、一定範囲内で0から1の実数値として算出する。フレーム番号 a および b で表される2フレーム間の時間的類似 St は、式(4)のように求められる。 fps は動画のフレームレート、 R は正規化区間[秒]である。 St は $a = b$ の場合に最大値1、 a から b までの時間的距離 $(|a-b|/fps)$ が R [秒]のときに最小値0となる。

$$St = 1 - \frac{|a - b|}{fps \times R} \quad (4)$$

4. 性能評価

4.1 鮮明度計算

視覚評価部における鮮明度計算の妥当性を評価するため、標準画像データベース SIDBA (図3) の12画像にガウシアンフィルタを適用したブレ画像を作成し、鮮明度を計算した。



図3 SIDBA (標準画像)

ブレ画像生成時は、フィルタのカーネルサイズを縦方向、横方向に変化させ、複数の加工画像を生成した。一例として、「couple」でブレ画像を生成した結果の一部を図4に示す。同様に、縦方向にもブレ生成を行う。



図4 「couple」のブレ画像 (横方向)

これらの画像群に対し、縦横それぞれの鮮明度の変化をプロットしたグラフを図5に示す。縦横で差はあるものの、カーネルサイズが大きくなるにつれて鮮明度が単純減少となっており、鮮明度をブレの指標として用いるのに適当であるといえる。

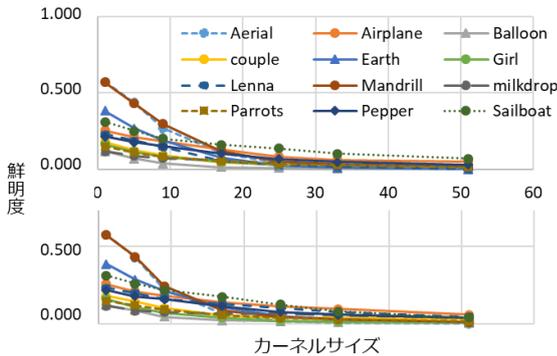


図5 ブレと鮮明度の変化 (上:縦方向, 下:横方向)

4.2 視覚的類似度計算

類似画像除外部における視覚的類似度計算の妥当性を評価するため、図3の画像から疑似的な類似画像を生成し、類似度を計算した。元画像の上下左右から1/3、1/2、2/3を白で塗りつぶした画像を用いる。1/3を塗りつぶした画像は類似度0.67、1/2では類似度0.50、2/3では類似度0.33となれば妥当と考える。生

成した類似画像の一例(上を塗りつぶした場合)を図6に示す。この画像群に対して式(2)で 4^3 , 8^3 , 16^3 , 32^3 , 64^3 色に減色し、式(3)で元画像との視覚的類似度を算出した結果を図7に示す。

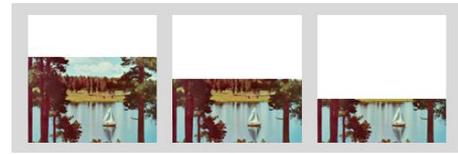


図6 類似画像の生成 (上を塗りつぶした場合)

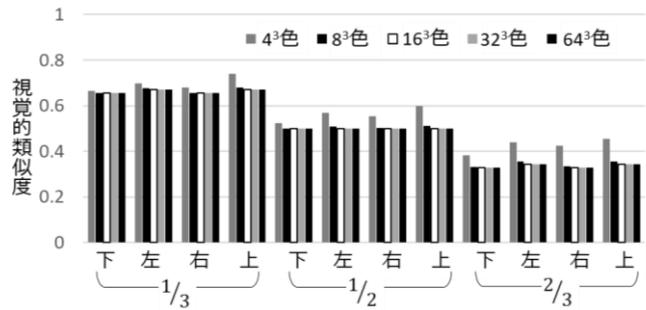


図7 類似度計算結果

この結果から、 4^3 色での結果が概して悪く、それ以外の減色数では正しく算出できていることが分かる。これは、減色後の色数が少ないために、塗りつぶす領域に関して、減色後の元画像と塗りつぶした色が一致する確率が高くなることの原因であると考えられる。そのため、計算方法としては妥当であると考えられる。

5. おわりに

本稿では、動画情報から利用者が思い出を振り返る際に楽しさを感じる場面の自動判定を行う手法のうち、特に「視覚評価部」と「類似画像除外部」について述べ、標準画像データベース SIDBA を用いて計算の妥当性評価を行った。鮮明度計算においては、ブレが大きくなるにつれて鮮明度が単純減少となり、ブレの指標として鮮明度が適当であることが示された。視覚的類似度においては、 4^3 色に減色した場合の結果が悪かったが他の減色数では正しく算出できており、計算方法として妥当であることが示された。今後は、視覚評価部および類似画像除外部のユーザ評価およびシステム全体の有用性について評価を行う。

謝辞

本研究の一部は、日本学術振興会 学術研究助成基金助成金の支援を受けたものである。

参考文献

- [1] 堀, 相澤; “ライフログビデオのためのコンテキスト推定”, 信学技報 IE, 画像工学, 103.514, pp.67-72, 2003
- [2] 中村; “映像によるライフログ”, 情報の科学と技術, pp.57-62, 2013
- [3] 木下, 小坂, 藤波; “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”, 情処研報, Vol.2016-UBI-51, No. 7, 2016.
- [4] 木下, 藤波; “身体装着型カメラの映像を用いた集団活動時の楽しい振り返りのための体験自動記録—会話場面の判定および評価—”, 情処学会第79回全国大会講演論文集, 2017(1), pp.511-512, 2017
- [5] 水上, 西村; “デジタル画像の色彩とエン트로ピー.” Memoirs of Faculty of Education, Shiga University 65 : pp.55-69, 2015