

K-059

ユーザ体験の時空間サンプリングによる ライフログ映像のキーフレームの抽出

Key-Frame Extraction for Life Log Video,
Using Spatio-Temporal Sampling of User's Experiences

河崎 晋也†
Shinya Kawasaki

石川 尊之‡
Takayuki Ishikawa

山崎俊彦‡
Toshihiko Yamasaki

相澤 清晴‡
Kiyoharu Aizawa

1. まえがき

近年、個人体験の記録の自動化が注目されている [1-4]。また、ウェアラブルコンピュータは、常に使用者の傍らにあるため、使用者と同じ状況下におかれる。この特徴は、個人の体験記録取得に於いて非常に有利である。この点に着目し、我々は手法としてウェアラブルコンピュータを用い、個人の体験記録を取得するというライフログシステムの開発に取り組んでいる[3], [4]。

一般の映像と比較すると、体験映像には非常に膨大な映像であり、かつ大部分は必要とされない画像であるという特徴がある。そのため、体験映像の閲覧及び編集の際には、求める映像を膨大な体験映像の中から効率よく検索することが必要とされる。映像へのインデキシングを行うことで検索速度の高速化を行うことができる。これまで我々は、ウェアラブルセンサーを用いて、映像と同時にコンテキストを表す特徴量を取得しインデキシングを行ってきた。より柔軟で効率のよい検索には、コンテキストに加えてコンテンツの利用が考えられる。本論文では、コンテキストとして時空間に着目したキーフレーム抽出を紹介する。

2. ライフログシステム

図 1 の様に我々は、撮影者個人の情報や撮影時の環境情報といった特徴量を、GPS センサや加速度センサ等のウェアラブルセンサーから取得している。また図 2 に示すように撮影した映像やセンサ情報は、ハードディスクに記録され、同期がとられている。また、インターフェイスでは、地図上での指定等、直感的な操作をコンテキストに基づいて行うことが可能となっている。

3. コンテキストとコンテンツの利用

3.1 コンテンツからコンテキストの抽出

複数のコンテキストを組み合わせた推定を行うことで、抽象度の高い検索を実行が期待される。そのために、更なるコンテキストを現す特徴量として画像特徴量の利用が考えられる。画像のようなコンテンツを解析することで輝度などのコンテキストを新たに得ることが可能となる。しかしながら、このような画像処理には処理が重いという問題がある。一方ライフログ映像には、そのほとんどが必要とされない映像であるという特徴がある。そこで画像処理の前処理として、コンテキストに基づいて映像からキーフレームを抽出する。これによって必要と

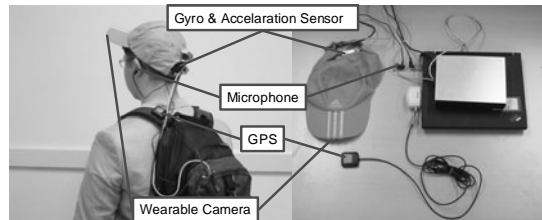


Figure 1. Wearable devices for Life Log.

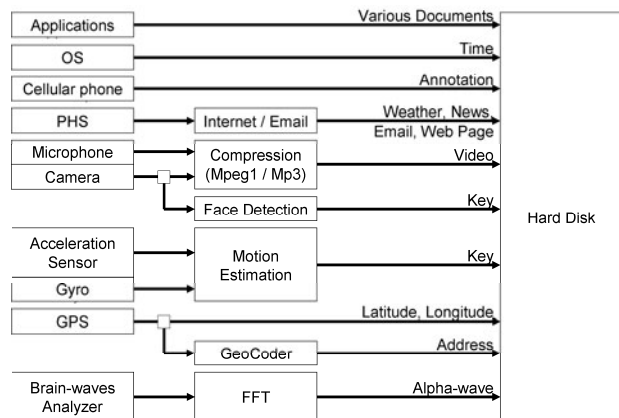


Figure 2. Block diagram of various types of data.

される映像だけを対象にすることが出来、効率的なコンテンツの解析が期待される。

3.2 コンテキストに基づいたコンテンツの表示

抽象度の高いクエリによる検索では、抽出結果も幅広いものとなる。そのため抽出結果を評価して、適切なクエリを導き出すことも必要である。抽出結果の表示法として、地図や年表のようなメタファを用いたコンテキストの表示は一目で概要が分かる点で有効であるが、前提としてメタファが一般的である必要がある。そのようなメタファを持たないコンテキストに対しては、一見して概要を認識できるフレームの一覧表示が有効と考えられる。

コンテキストに基づいて様々なフレーム抽出を行うことで、コンテキストを反映したコンテンツの表示ができ、効率的な抽出結果の評価が期待される。

4. 時空間サンプリング

コンテキストの違いによるキーフレームの変化を調べるために実験を行った。キーフレームの抽出実験として、日々蓄積した下校時の体験映像を対象にした。移動は自転車を利用している。コンテキストとしては、人間の回想に於いて時空間情報が重要な検索キーとなっていることに着目し、時刻と GPS による位置情報から抽出を行った。

†東京大学大学院工学系研究科

‡東京大学大学院新領域創成科学研究科



Figure 3. Extracted frames every 30 seconds.



Figure 4. Frames extracted every 50 meters.



Figure 5. Example of extracted frames when the speed or direction is changed.

4.1 均一な時間でのサンプリング

図 3 は 30 秒ごとに切り出した結果である。この映像が登校に関連する映像であるなど、一定時間でのサンプリングでも映像の概要を知ることが出来る。しかしながら、距離など登校路に関する情報はわからない。

4.2 均一な距離でのサンプリング

図 4 は 50m ごとに切り出した結果である。図 3 と比較して、均一時間で抽出と異なり、均一距離での抽出には撮影時の速度の変化に影響されない特徴があることが分かる。

4.3 速度での制御

速度によってサンプリングを制御した場合も、異なる抽出結果が得られた。速度が時速 4Km 以下であるときの抽出結果からは、交差点での停止時や自転車を降りたときなどの映像が抽出された。また、速度の変化をトリガとした場合には、発進・停止の外に走行中のスピードの変化時の映像も抽出されている。

4.4 方向での制御

進行方向が変化をトリガとして抽出を行うと、交差点などでカーブ時や自転車を停車したときの映像が抽出された。

4.5 複数のコンテキストによる制御

これらのコンテキストを組み合わせることで検索状況に適したサンプリングが期待される。例えばナビゲーションを念頭に置いた場合、発進・停止・カーブが重要である。そこで、速度変化と進行方向の変化を組み合わせで抽出を行った(図 5)。発進・停止・カーブ時のフレームが多く抽出されている。

5. まとめ

ライフログシステムにおいては、短時間での検索のために、自動的にインデックスをつけることが必要である。自動的なタグ付けは、低レベルの特徴量での検索が必要になるが一般にクエリが分かりにくい問題がある。しかし、コンテキストを反映した特徴量を用いれば直感的に分かり易く、かつ短時間での検索が可能となる。

そこで、これまでウェアラブルコンピュータを用いてコンテキストに基づく特徴量を取得し、それを用いて短時間且つ、分かりやすい検索を実現してきた。

今後の発展としてコンテキストの情報を組み合わせてライフログを分類し、より抽象的なクエリに対応できるようにすることが考えられる。その場合、コンテンツをあわせて利用することは重要であると考えられる。

そのほかに、日常行動か特異な行動であるかの判断や、屋内での位置検出を利用していく[5]ことも必要である。

6. 参考文献

[1] J. Gemmell, G. Bell, R. Lueder, S. Drucker and C. Wong, "MyLifeBits: Fulfilling the Memex Vision," ACM Multimedia System Journal, pp.235-238, 2002.

[2] 上岡, 広田, 広瀬, "ウェアラブルコンピュータによる主観的体験とその展開", ヒューマンインターフェース学会研究報告集, vol.5 No.4, 2003.

[3] 石島, 相澤, "個人体験映像の構造化と要約-生態情報を用いた映像要約によるライフメディア", 信学技報, IE2000-23, July, 2000.

[4] 堀鉄郎, 河崎晋也, 石川尊之, 相澤清晴, "ライフログデータの効率的検索," 電子情報通信学会 2004 年総合大会, 東京, Mar.22-25, 2004.