

## 手話単語の意味的特徴点と手話動画像圧縮への応用

1 L-5

宮尾 淳一

広島大学 総合科学部  
〒739-8521 東広島市鏡山 1-7-1  
miyao@mis.hiroshima-u.ac.jp

### 1 まえがき

筆者らはコンピュータによって手話学習を援助するシステムの研究を行っており、現在、手話単語特徴に基づく教授法およびそれを用いた手話学習システムSLCの開発を行っている。<sup>[1, 2]</sup> この教授法は、手話単語動作を時間、空間について詳細に解析し、認識に重要な要素の重み付けを行った手話単語特徴を用いて、キーポイントを重点的に教授し、かつ、学習者からのフィードバックもこれにより評価して矯正指導するものである。

この手話単語特徴の時間軸上の解析において、単語認識に最低限必要なキーとなる動作は単語により異なるが、1箇所から数ヶ所程度であることが分かった。そこで、このキーとなる動作を手話単語の意味的特徴点と呼び、各特徴点の静止画像のみを動画像として表示すると、ギクシャクするが十分認識できることが分かった。

本稿では、この意味的特徴点とそれを用いた超低ビットレートのための手話動画像の圧縮について述べる。

### 2 意味的特徴点

手話学習システムSLCにおける手話単語特徴は、動作の時間軸上の重要度を表す時間特徴と空間上の位置や形状の重要度を表す空間特徴に分けられる。筆者らは、CGを用いて手話動画像を合成し、それを用いて24単語の手話特徴を解析している。

時間特徴は、まず、モーションキャプチャによりCG合成した手話動画像(30フレーム/秒)の各静止画において動作の瞬間的停止点と動き方向の急変化点およびそれらの中間点を意味的特徴点の候補集合として抜き出す。次に、候補集合の要素を表示あるいは非表示にするすべての組み合わせに対して動画像を合成して、それらを人間が評価し最適な組み合わせを意味的特徴点として抽出したものである。

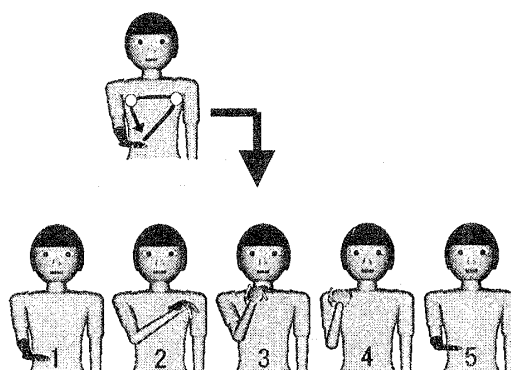


図 1: 単語「大丈夫」の候補集合

表 1: 抽出した単語と意味的特徴点数

手話単語	特徴点数	手話単語	特徴点数
あなた	1	行く	2
イメージ	4	嘘	1
覚える	2	思う	1
嫌い	2	困る	3
好き	2	すみません	2
大丈夫	2	尋ねる	2
誰	3	電話	1
ときどき	3	トマト	3
何	3	なるほど	3
本当	3	良い	2
分からない	3	分かる	2
忘れる	2	悪い	2

例えば、単語「大丈夫」の意味的特徴点の候補集合は図1のように5つの画像(調動開始終了点を含む)となるが、意味的特徴点は画像番号2と4の2点の画像となる。表1に、筆者らが分類した各動作形状を含み日常よく用いると考えられる24単語とそれらの特徴点数を示す。

### 3 準意味的特徴点

意味的特徴点は品質を問わず最低限の情報伝達を目的とした手話動画像の圧縮に応用できると考えられるが、実際に応用するためには画像からの自動的な手話

単語認識が必要となり、実時間での処理には困難が伴う。そこで、形式的に抽出する方法が必要となる。

意味的特徴点は手話調動において、動作が一瞬止まった点や急に動作方向が変わった点、および、それらの中間点の候補集合の部分集合により構成されている。そこで、候補集合あるいはそれを絞込んだ集合を準意味的特徴点として抽出することにする。

実際に準意味的特徴点をビデオ画像から検出するプログラムを画像間の輝度差分を基にして作成しパラメータを調整したところ、次節でも述べるように不要な画像は比較的少ない良好な抽出結果が得られた。

また、この準意味的特徴点を検出するための画像間差分や動きベクトル推定は既存のH263<sup>13)</sup>などにおいて既に組み込まれており、既存方式に容易に組み込むことができ、かつ、実時間処理も期待できる。

#### 4 実写画像による抽出実験

手話動画を実際にビデオ撮影し、その各フレームから意味的特徴点および準意味的特徴点を抽出する実験を行った。

まず、ビデオ画像は本学聾学校教員養成課程の学生2名にお願いし、3つの短文について表現してもらった。それをデジタルビデオカメラ(SONY DCD TRV-2000)により撮影し、IEEE1394のDV端子からコンピュータに読み込み、640×480ドットの静止画像列として利用した。圧縮を行う前の原画像のフレーム数は表2のようになった。

次に、人手で原画像から意味的特徴点を抜き出し、これにより動画を構成したところ、表示はギクシャクするが、意味は完全に伝達されることを確認した。また、この意味的特徴点を構成する静止画像の枚数を表2に示す。続いて、原画像より準意味的特徴点を抽出するプログラムにより自動的に抽出した静止画像の枚数は表2のようになった。

比較対象として画像の等間隔間引きによる圧縮を行った。今回の実験では全単語が認識できる最大の間引き間隔を求めた。ただし、間引きをどの画像から開始するかによっても認識結果が異なるので、必ず全単語が認識できる最大の間隔を選んだ。実験結果を表2に示す。

以上の実験より、意味的特徴点による手話動画の圧縮は原画像により変化することは少なく約1/7になることが分かる。これを達成するためには手話認識が必要となるが、形式的に検出できる準意味的特徴点による圧縮でも約1/6程度に圧縮できることが分かった。

また、比較対象とした等間隔間引きによる圧縮では

表 2: 実写画像から各方式で抽出した画像

手話動作	時間 (秒)	フレーム数			
		圧縮前	意味的 圧縮	準意味 的圧縮	等間隔 圧縮
電話忘れてすみませんF	4.0	60	7	10	12
電話忘れてすみませんG	4.3	65	7	9	10.8
私はトマトが嫌いですF	4.3	65	8	10	10.8
私はトマトが嫌いですG	4.3	65	8	8	12
あなたは何が好きですかF	4.3	65	10	12	32.5
あなたは何が好きですかG	4.3	65	10	10	32.5

準意味的特徴点と同様な結果となる場合もあるが、それより3倍程度悪い場合もあり、等間隔間引きによる方法は伝達する手話画像により最適な間引き間隔が変動することが分かった。この等間隔圧縮を実際に応用するためには最悪の場合を想定して間引き間隔を設定する必要があるため、全体的な圧縮率は約1/2程度になると予想される。

#### 5 あとがき

本稿では、手話動画の意味的特徴点とそれの手話動画圧縮への応用を考察した。意味的特徴点をそのまま実際に応用することは困難であるが、自動的に抽出できる準意味的特徴点を用いることにより、意味的特徴点と同様な圧縮が行えることを示した。また、等間隔で間引く方法と比較して、その半分以下に圧縮できる可能性を示した。

今後は、準意味的特徴点の抽出アルゴリズムの改良や実時間処理の検討などを計画している。

#### 謝辞

今回の実験に協力して頂いた本学学校教育学部聾学校教員養成課程の学生さん、および、プログラム作成、実験などに協力して頂いた本学総合科学部生福島純一氏(現、長野日本ソフトウェア(株))にも感謝致します。

なお、本研究の一部は文部省科学研究費補助金基盤研究(C)(2)(課題番号 09680223)によるものである。

#### 参考文献

- [1] 宮尾, 中間: “手話学習システムのための手話表現基礎データの作成”, 信学技報, ET97-123, pp.81-88, 1998.
- [2] 宮尾, 中間: “手話単語特徴に基づく教授法と手話学習システム”, 信学技報, ET98-130, pp.89-92, 1999.
- [3] K.R. Rao and J.J. Hwang: “Techniques & Standards for Image, Video & Audio Coding”, Prentice-Hall, 1996.