

動きの向きの情報を持つモーションヒストリーイメージによる 人の動作の表現と認識

緒方 健人[†] タン ジュークイ[†] 金 亨燮[‡] 石川 聖二[†]
^{†‡} 九州工業大学

E-mail:†{ogata, etheltan, ishikawa}@ss10.cntl.kyutech.ac.jp, ‡kim@cntl.kyutech.ac.jp

あらまし 本稿では、動きの向きの情報を持つモーションヒストリーイメージによる人の動作の表現法及び認識法を提案する。既存手法であるモーションヒストリーイメージは時空間画像で表現される人の動作を2次元画像で表現する手法であるが、3次元情報を圧縮するため、複雑な動きを表現する場合に新しい動きが古い動きを上書きするという問題がある。提案手法では、オプティカルフローから得られる動きベクトルを4つの成分に分解してモーションヒストリーイメージを生成することによって方向性を付加し、この問題の解決を図る。動作認識にはHuモーメントに基づく特徴量を用い、大きな動きが描かれた方向に重みを置く認識方法を用いる。実験では、同じ実験データに対し従来手法と提案手法でそれぞれ認識率を求め、提案手法の有効性を示す。

キーワード 動作表現, 動作認識, モーションヒストリーイメージ, オプティカルフロー.

Representation and Recognition of Human motions by Directional Motion History Images

Takehito OGATA[†], Joo Kooi TAN[†], Hyoungseop KIM[‡], and Seiji ISHIKAWA[†]
^{†‡} Kyushu Institute of Technology

E-mail:†{ogata, etheltan, ishikawa}@ss10.cntl.kyutech.ac.jp, ‡kim@cntl.kyutech.ac.jp

Abstract: A new method for human motion representation and recognition is proposed. The motion history image proposed by Bobick et al. can represent human motion by a single 2-D image. However due to its simple representation, this image sometimes erases its old motion by overwriting a new motion. To avoid this situation, four individual motion history images are generated from directionally separated optical flow components. Hu moment-based features are used for recognition like Bobick's method, however importance among four motion history images is considered. In order to evaluate the performance of the proposed method, its results are compared with those obtained by the original method.

Key words: Motion representation, Motion recognition, Motion history images, optical flow.

1. はじめに

コンピュータが人の動作を認識するためには人の動きを取得しなければならない。その場に置くだけで空間情報を非接触で画像として取得できる視覚センサは、対象者がセンサなどを装着する必要がないため、対象者が撮影されていることを意識していない動作を認識することができる。このような動作を認識する場合、対象者はカメラに大きく写っているとは限らないため、対象者の解像度が低い場合にも有効な動作認識手法が必要である。

視覚に基づく人の動作に関する研究は非常に幅広く、様々な基準で分類が可能であるが、画像内の人の解像

度を基準にすると3種類に分類できる[1]。手話データの解析など人の解像度が大きい場合、人体モデルに基づく手法が主流であり、体の各パーツを個々に検出し姿勢を推定する手法[2]や複数台のカメラを用いて3次元復元する手法[3]などが提案されている。逆に、上空から撮影された映像など人の解像度が数ピクセルの場合や体の動き情報が少ない場合は、人を追跡しその軌跡や速度などを解析する手法が用いられる[4]。そして人の体の動きが見えるが体の各部位の詳細までは得られない中程度の解像度の場合、体の姿勢の変化や動き情報をそのまま解析する、見え方に基づく手法が主流である。既存の研究をこのように分類した場合、一般に低い解像度で開発された手法は高い解像度に適用で

きるが、その逆は困難であることが言える。解像度が低い映像を対象とした人のモデルに基づく手法も提案されている[5,6]が、一般に人のモデルに基づく手法は人の体の自由度が大きいためこれらの手法は常に計算コストの問題を持つ。また低解像度の手法は体の動きが得られない場合の手法であり、体の動きが得られる場合はその動きをみる手法が望ましい。よって、われわれは見え方に基づく動作認識手法に着目する。

見え方に基づく動作認識として、Bobick らのテンポラルテンプレートを用いた手法がある[7]。テンポラルテンプレートの一つであるモーションヒストリーイメージ(MHI; Motion History Image))はその画像中に過去の情報を輝度情報として埋め込むため、時空間で表現される人の動作を 2 次元の画像として表現できる。しかし MHI は時間情報を輝度値として表現するため、複雑な動作を表現する場合に、同じ位置に複数回動きが発生すると古い動きが新しい動きによって上書きされ失われるという欠点がある。

本稿では、動きの向きの情報を持つ方向性モーションヒストリーイメージを用いた動作表現法および認識法を提案する。従来の MHI の問題点である上書き問題を解決するため、オプティカルフローから得られる動きベクトルを上下左右の 4 つの成分に分解し、これらから独立してモーションヒストリーイメージを生成する。動作認識には Bobick らの手法と同様に Hu モーメントに基づく特徴量を用いるが、モーションヒストリーイメージ間の関係を考慮し大きな動きが描かれた方向に重みを置く認識方法を用いる。実験では動きの向き情報の有効性を検討するため、同じ実験データに対し Bobick らの従来手法と提案手法を適用しそれぞれ認識率を求め比較を行う。

本稿の構成は、2 節にて方向性モーションヒストリーイメージによる動作の表現方法、3 節にてこれを用いた動作認識方法について述べ、4 節にて実験結果を示し、最後に結論を述べる。

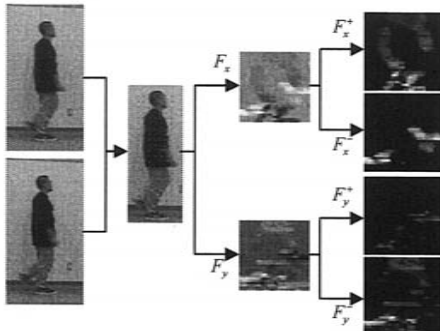


図 1. オプティカルフローの分解の様子

2. 方向性モーションヒストリーイメージによる動作の表現

2.1 動きの向き情報の導入

モーションヒストリーイメージ(MHI; Motion History Image)は一つの画像に過去の画像を残していく時系列画像の表現方法の一つである[7]。時系列画像から MHI を生成すると、新しい画像ほど明るく、過去の画像ほど暗く表示される。Bobick らにより提案された MHI $H_\tau(x, y, t)$ は時系列の 2 値画像から、次式により生成される。

$$H_\tau(x, y, t) = \begin{cases} \tau & \text{if } D(x, y, t) = 1 \\ \max(0, H_\tau(x, y, t-1)) - 1 & \text{otherwise.} \end{cases} \quad (1)$$

ここで、 x, y, t はそれぞれ画像の画素の座標値および時間を示し、 $D(x, y, t)$ は 2 値入力画像、 τ はどれだけの期間の画像を残すかを決定させるパラメータである。

これに対応する、 τ の期間に動きのあった領域を表現する 2 値画像モーションエナジーイメージ(MEI; Motion Energy Image)は次式で与えられる。

$$E_\tau(x, y, t) = \begin{cases} 1 & \text{if } H_\tau(x, y, t) > 0 \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

一方、提案法はオプティカルフローを 4 つに分解した成分から MHI を生成する。オプティカルフローの分解方法は Efros らのモーションディスクリプタ(MD; Motion Descriptors)と同じであり、オプティカルフローを水平成分 F_x 及び垂直成分 F_y に分解、さらに F_x 及び F_y をそれぞれ正の成分 F_x^+ 、 F_x^- 及び負の成分 F_x^- 、 F_y^- に分解する。この分解の様子を図 1 に示す。分解して得られる 4 つの成分は、画像上における動きの方向を表す。一般的な画像座標系の場合、 F_x^+ 、 F_x^- 、 F_y^+ 、及び F_y^- はそれぞれ右、左、下、及び上方向に対応する。以降、これらの成分を方向成分と呼ぶ。

分解により得られる 4 つの方向成分を用いて、以下の式により 4 つの独立した MHI を生成する。

$$H_\tau^{+x}(x, y, t) = \begin{cases} \tau & \text{if } F_x^+(x, y, t) > th_H \\ \max(0, H_\tau^{+x}(x, y, t-1)) - 1 & \text{otherwise} \end{cases} \quad (3)$$

$$H_\tau^{-x}(x, y, t) = \begin{cases} \tau & \text{if } F_x^-(x, y, t) > th_H \\ \max(0, H_\tau^{-x}(x, y, t-1)) - 1 & \text{otherwise} \end{cases} \quad (4)$$

$$H_\tau^{+y}(x, y, t) = \begin{cases} \tau & \text{if } F_y^+(x, y, t) > th_H \\ \max(0, H_\tau^{+y}(x, y, t-1)) - 1 & \text{otherwise} \end{cases} \quad (5)$$

$$H_\tau^{-y}(x, y, t) = \begin{cases} \tau & \text{if } F_y^-(x, y, t) > th_H \\ \max(0, H_\tau^{-y}(x, y, t-1)) - 1 & \text{otherwise} \end{cases} \quad (6)$$

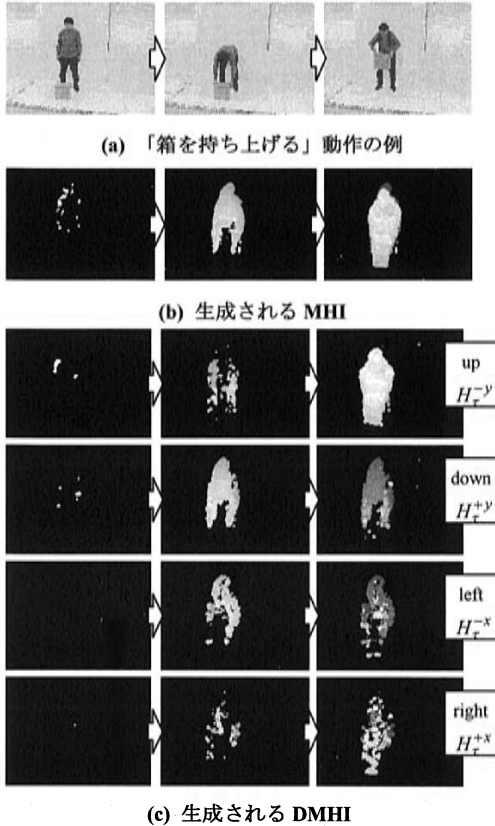


図2. 上書き問題の軽減の例

ここで $H_t^{+x}(x,y,t)$, $H_t^{-x}(x,y,t)$ は左右の動きから生成される MHI, $H_t^{+y}(x,y,t)$, $H_t^{-y}(x,y,t)$ は上下の動きから生成される MHI であり, th_H は方向成分の閾値である. この閾値処理は, 式(4.1)の MHI の 2 値入力画像 $D(x,y,t)$ を生成することに対応する. 式(4.3)~(4.6)より生成される MHI は, それぞれ上下左右の方向に発生した動きの履歴を表している. よって, 我々はこれらの MHI を動きの向きの情報を持つ方向性モーションヒストリーイメージ(DMHI; Directional Motion History Images)と呼ぶ. 本稿では, 人の動作を 4 つの DMHI のセットで表現する.

2.2 上書き問題の解決

動きの向きに分解された成分からそれぞれ独立した MHI を生成することにより, 従来の MHI では動きの向きの変化によって上書きされ失われていた過去の動き情報を残すことができる. 図2にその例を示す. 同図は「箱をつかんで持ち上げる」という動作から MHI を生成したものの一部で, (a)は原画像, (b)は従来の

MHI, (c)の4つの画像は上から順に, それぞれ F_y^- , F_y^+ , F_x^- および F_x^+ から生成した DMHI である. この動作は, 最初人が箱をつかむ際に体を下にかがめ, 箱をつかんだ後上体を上げるという動作であり, 体をかがめた時点のフレームが図の2列目, 箱をつかんで上体を上げた時点のフレームが3列目に示されている. ここで従来の MHI を見ると2列目の時点で描かれている上体を下げるという動きの情報が, 3列目の上体を上げるという動きの情報によって上書きされ, 情報が失われていることがわかる. 一方, DMHI を見ると, 2列目における上体を下げるという動きは下方向の成分 H_t^{+y} に描かれ, 3列目における上体を上げるという動きは上方向の成分 H_t^{-y} に描かれているため, 上書きが回避され最初の下方向の動きの履歴が失われず残っていることがわかる.

3. 動作認識法

Bobick らの動作認識手法は位置, 大きさ, 回転に不変な特徴量である Hu モーメント[8]に基づいており, MHI 及び MEI それぞれから得られる計 14 個の特徴量を認識に用いる. この Hu モーメントに基づく動作認識手法を DMHI に応用した動作認識手法を以下に示す.

まず, MHI の場合と同様に, 各方向の DMHI から動きの情報を持つモーションエナジーイメージ(DMEI; Directional Motion Energy Images)をそれぞれ生成する. DMEI は従来の MEI の生成式(2)を DMHI に適用することにより得られる. すなわち, DMHI H_t^{+x} , H_t^{-x} , H_t^{+y} 及び H_t^{-y} を閾値 0 で 2 値化することにより, 対応する DMEI E_t^{+x} , E_t^{-x} , E_t^{+y} 及び E_t^{-y} が生成される. 各 DMHI とそれに対応する DMEI から, 7 種類の Hu moment をそれぞれ求め, 合計 14 個の特徴量を 4 組生成する.

生成された 4 組の特徴量は, それぞれの方向の DMHI 及び DMEI の形状情報を持つが, どの方向の DMHI により多くの動き情報が含まれるか等の, DMHI 相互の関係の情報は持たない. そこで面積に着目し, DMHI の面積の割合及び DMEI の面積の割合の 2 種類の特徴量を, 対応する方向の特徴量の組に追加する. 面積には 0 次モーメントを用い, DMHI の 0 次モーメントは 4 つの DMHI の 0 次モーメントの合計で, また DMEI の 0 次モーメントは 4 つの DMEI の 0 次モーメントの合計で除算することにより正規化する. DMHI H_t^{+x} , H_t^{-x} , H_t^{+y} 及び H_t^{-y} の 0 次モーメントがそれぞれ $m_{00}(H_t^{+x})$, $m_{00}(H_t^{-x})$, $m_{00}(H_t^{+y})$ 及び $m_{00}(H_t^{-y})$ から求められるとすれば, 例えば DMHI H_t^{+x} の正規化 0 次モーメント \hat{m}_H^{+x} は式

$$\hat{m}_H^{+x} = \frac{m_{00}(H_\tau^{+x})}{m_{00}(H_\tau^{+x}) + m_{00}(H_\tau^{-x}) + m_{00}(H_\tau^{+y}) + m_{00}(H_\tau^{-y})} \quad (7)$$

より算出され、残りの \hat{m}_H^{-x} , \hat{m}_H^{+y} 及び \hat{m}_H^{-y} も式(7)と同様の分母で正規化することで得られる。DMEI の正規化 0 次モーメント \hat{m}_E^{+x} , \hat{m}_E^{-x} , \hat{m}_E^{+y} 及び \hat{m}_E^{-y} も同様に全 DMEI の 0 次モーメントの合計で正規化することで算出される。こうして得られる各方向の 16 種類の特徴量をそれぞれ 16 次元の特徴ベクトルとして構成し、これら 4 組の 16 次元の特徴ベクトルを動作認識に用いる。

動作の認識は各方向で独立してパターン認識を行い、その結果を先に求めた DMEI の正規化 0 次モーメントを重みとした重みつき多数決で決定する。よって、あらかじめ学習段階で認識させたい動作の動画像から DMHI を生成し、前述の方法で 4 組の特徴ベクトルを算出し、これらを別々の 16 次元の特徴空間に投影し、それぞれに動作の名前を対応させておく。そして認識の段階で、学習に用いていない動作から同様に DMHI を生成し前述の方法で 4 組の特徴ベクトルを求め、各組で独立して特徴空間内で認識処理を行う。提案法では、この認識法として学習データのサンプル全点とのユークリッド距離を用いた探索による k -近傍法(k -NN)を用いる。全点探索の手法は計算コストが大きいという欠点はあるが、複雑な分布を持つクラスタの認識に適している[9]。認識の結果、学習段階であらかじめ用意していた動作のうちどの動作に近いかという結果が 4 組得られる。

最後に、得られた 4 つの認識結果を重みつき多数決により統合する。重みには、対応する方向の DMEI の正規化 0 次モーメント \hat{m}_E^{+x} , \hat{m}_E^{-x} , \hat{m}_E^{+y} 及び \hat{m}_E^{-y} を用いる。DMHI ではなく DMEI の正規化 0 次モーメントを用いるのは、動きの新しさ、古さに関係なく、単純により広い領域に動きが描かれる方向の DMHI が重要な情報を持っていると考えるためである。よって、DMHI H_τ^{+x} , H_τ^{-x} , H_τ^{+y} 及び H_τ^{-y} の認識結果をそれぞれ g_τ^{+x} , g_τ^{-x} , g_τ^{+y} 及び g_τ^{-y} とすると、重みつき多数決による最終的な認識結果 g_F は次式より得られる。

$$g_F = \arg \max_{a^* \in A} \left(\hat{m}_\tau^{+x} \delta(g_\tau^{+x}, a^*) + \hat{m}_\tau^{-x} \delta(g_\tau^{-x}, a^*) + \hat{m}_\tau^{+y} \delta(g_\tau^{+y}, a^*) + \hat{m}_\tau^{-y} \delta(g_\tau^{-y}, a^*) \right) \quad (8)$$

ここで a^* は動作の種類、 A は学習段階で用意した動作の集合、 $\delta(g, a)$ は認識結果 g と動作 a が同じ場合に 1、異なる場合に 0 を出力するようなデルタ関数である。この動作認識処理の概略を図 3 に示す。

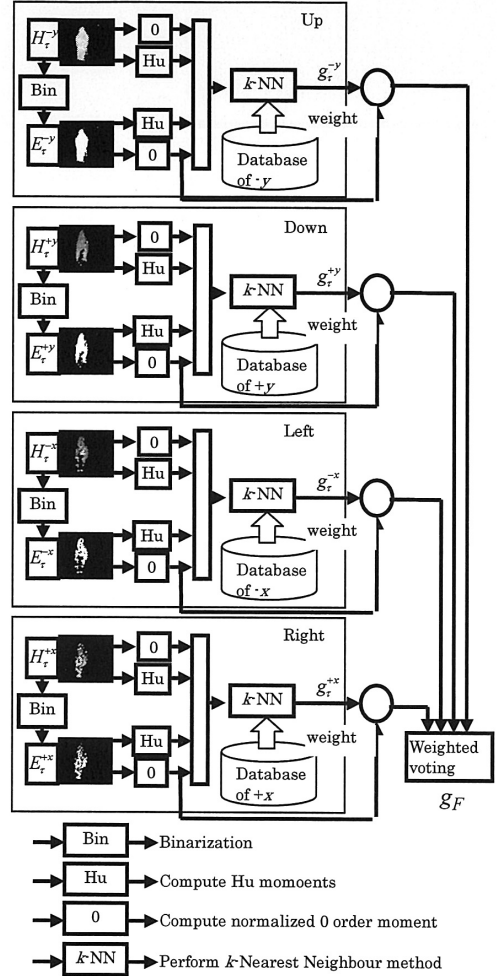


図 3. 認識処理の概略

4. 認識実験

提案した手法を用いて実験を行った。性能の評価のため、同じ実験データに対して従来手法と提案手法を適用し認識率の比較を行った。

4.1 実験環境

実験データは対象者の向いている方向を中心に取り囲むように配置された 4 台のカメラを用いて、11 人の 5 種類の動作を撮影した。動作は「箱を持ち上げる」、「小さなものを拾う」、「うずくまる」、「その場で歩く」、そして「倒れる」である。動作の例を図 4 (a) に示す。

動作 1 つを 1 つの動画ファイルとして切り出し、その動画ファイルの最終フレーム時点で得られる MHI 及び DMHI を認識に用いた。

表 4.1 実験結果

動作	認識率[%]	
	(a) MHI	(b) DMHI
箱を持ち上げる	54.6	100
小さなものを拾う	63.6	79.6
うづくまる	84.1	88.6
その場で歩く	81.8	100
倒れる	63.6	97.7
平均	69.6	93.2

実験の学習には Leave-one-out 法を用いた。一人分のデータをテスト用に取り出し、残りのデータを用いて学習を行った。学習時サンプルをデータベースに登録する際にはカメラの向きに関係なく動作ラベルのみを登録した。

パラメータの設定は以下の通りである。MHI 及び DMHI に残す履歴の長さを決定するパラメータ τ は各動作のフレーム数と同じとし、MHI の入力画像を生成するためのフレーム間差分画像の閾値は 40、DMHI の閾値 th_H は 1.0、認識のための k -NN は $k=1$ とした。なお、従来法は $k=1$ 及び 3 で 2 回 k -NN 法による実験を行い、より認識率の高かった $k=3$ を採用した。

4.2 実験結果

実験結果より得られた認識率を表 1 に示す。表 1 (a) が従来法の結果、(b) が提案手法の結果である。また、この実験で得られた MHI 及び DMHI の例を図 4(b) 及び(c)にそれぞれ示す。

5. 考察

実験結果より、動きの向きの情報を持つ DMHI は通常の MHI よりも全体の認識率が向上していることが確認できる。これは、2.2 節で述べたように、動きの向きを分けて MHI を生成することによって、異なる向きの動きによる過去の動きの上書きが発生しないことが有利に働いているためと考えられる。各実験の結果について以下で詳しく検討する。

まず、DMHI と MHI の性能比較を行うと、表より、「箱を持ち上げる」、「その場で歩く」、そして「倒れる」の場合、DMHI の認識率は大きく改善された一方、「小さなものを拾う」と「うづくまる」ではあまり改善されなかったことがわかる。この原因として、認識率があまり改善されなかった 2 つの動作は、どちらも下方への動きしか含まないような動作であったためと考えられる。提案手法はまず各方向の DMHI 及び DMEI による認識を独立して行うが、このうち、大きな動き

が描かれていない方向の DMHI 及び DMEI の認識結果の信頼性は低い。この問題を解決するため、提案手法は DMEI の正規化 0 次モーメントを重みとした重みつき多数決を採用している。しかし、図 4 (c)を見ると、認識率が改善されなかった 2 つの動作はほぼ下方方向の動きしかないはずであるが、他の 3 方向にも動きが描かれており、これが信頼性の低い方向の認識結果の重みを増加させ誤認識を招いたものと考えられる。他の方向に動きが描かれたのは、本実験で使用した勾配法に基づくオプティカルフロー算出手法の精度があまり良くなかったためと考えられる。この問題を改善するには、より信頼性の高いオプティカルフローの算出法を考案する必要がある。また、この結果より、DMHI は、複数の方向への動きを含む動作の認識に特に適していると言える。

提案手法のボトルネックは計算負荷である。従来手法は MHI をフレーム間差分から生成できるのに比べ、提案手法は DMHI を生成するためにオプティカルフローを計算する必要があるため従来法より計算負荷が大きい。しかし、これは最近のハイエンド PC を使用することで軽減することができる。筆者らの環境では、Pentium 4 3GHz、メモリ 2GB の PC で画像サイズ 160 × 120 の場合、30[fps]で動作することを確認している。また、オプティカルフローの計算自体は単純であり FPGA などで計算することができるため、このようなハードウェアを導入することによる改善も期待できる。

6. おわりに

本論文では動きの向きの情報を持つモーションヒストリーイメージを用いた人の動作の表現法及び認識法を提案した。従来法のモーションヒストリーイメージの動きの上書き問題を解決するため、Efros らの手法を参考にオプティカルフローの成分を 4 つに分割し、それぞれから独立してモーションヒストリーイメージを生成した。提案手法及び従来手法を同じ実験データに適用し、提案手法の認識性能が従来手法より良かったことより、動きの向きの情報を持つ DMHI の有効性が確認できた。また実験結果より、提案手法は複数の方向への動きを含む動作の認識に特に適していることが判明した。提案手法はオプティカルフローの計算を含むため計算負荷が大きいという弱点があるが、これはハイエンド PC や FPGA などのハードウェアを導入することによって克服できる。

提案手法は、より高精度、かつ計算コストの低いオプティカルフロー算出法の導入により更なる性能の向上が期待できる。

参考文献

- [1] A.A. Efros, A.C. Berg, G. Mori and J. Malik: "Recognizing action at a distance", *In Proc. of ICCV*, Vol. 2, pp. 726-733 (2003).
- [2] A.S. Micilotta and R. Bowden: "View-based location and tracking of body parts for visual interaction", *In Proc. of BMVC*, Vol. 2, pp. 849-858 (2004).
- [3] 伊達 直人, 吉松 寿人, 有田 大作, 谷口 倫一郎: "多視点動画画像解析による身体の実時間姿勢推定", 画像の認識・理解シンポジウム, Vol.2, pp.678-683 (2004).
- [4] 小島 篤博, 田原 典枝, 田村 武志, 福永 邦雄: "動画画像における人物行動の自然言語による説明の生成", 電子情報通信学会論文誌 D-II, Vol. J81-D-II, No.8, pp. 1867-1875 (1998).
- [5] D. Hogg: "Model-based vision: a program to see a walking person", *Image and Vision Computing*, Vol.1, No.1, pp.5-20 (1983).
- [6] H. Sidenbladh, M.J. Black, and L. Sigal: "Implicit probabilistic models of human motion for synthesis and tracking", *In Proc. of ECCV*, Vol. 1, pp.784-800 (2002).
- [7] A.F. Bobick, J.W. Davis: "The recognition of human movement using temporal templates", *IEEE Trans. on PAMI*, Vol. 23, No. 3, pp.257-267 (2001).
- [8] M. Hu: "Visual pattern recognition by moment invariants", *IRE Transactions on Information Theory*, Vol. 8, No. 2, pp.179-187 (1962).
- [9] 甘利俊一, 麻生英樹, 津田宏治, 村田昇: パターン認識と学習の統計学—新しい概念と手法, 岩波書店 (2003).

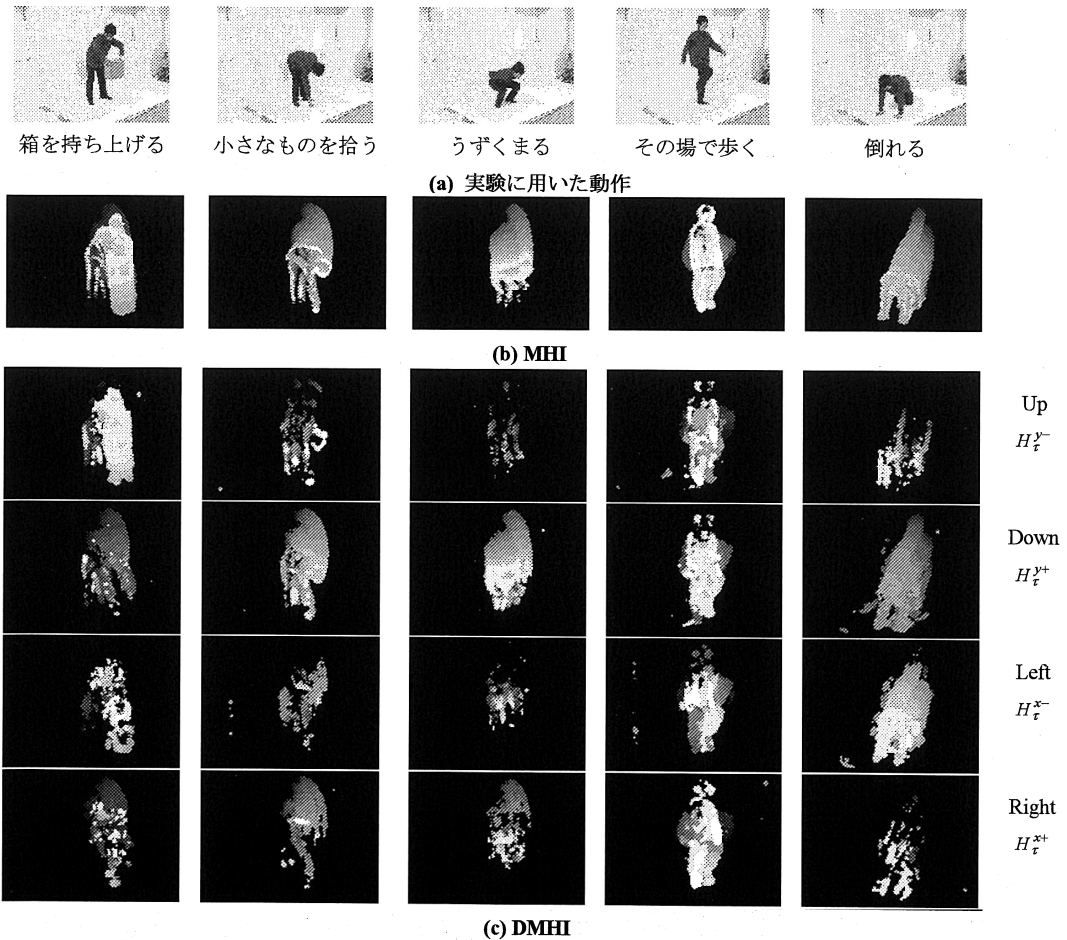


図 4. 実験に用いた各動作から生成される MHI 及び DMHI の例