高次局所自己相関特徴量に基づく 1000fps 高速画像認識

1000 fps High-Speed Image Recognition based on HLAC Features

資延 亮 (広島大学)
高木 健 (広島大学)山本 健吉 (広島大学)
石井 抱 (広島大学)Ryo Sukenobe, Hiroshima University, sukenobe@robotics.hiroshima-u.ac.jp
Kenkichi Yamamoto, Hiroshima University, yamamoto@robotics.hiroshima-u.ac.jp
Takeshi Takaki, Hiroshima University, takaki@robotics.hiroshima-u.ac.jp
Idaku Ishii, Hiroshima University, iishii@robotics.hiroshima-u.ac.jp

Real-time image recognition at 1000 fps is realized by implementing a circuit module to calculate higher-order local auto-correlation (HLAC) features on a high-speed vision platform. The high-speed vision platform, on which the HLAC circuit module is hardware-implemented, can extract 25 HLAC features at 1000 fps from 1024×1024 pixel images. In the experimental results, high-speed projected images are recognized by using HLAC features.

Key Words: High-speed Vision, Robot Vision, Real-time Sensing, Image Recognition, Integrated Sensing

1. はじめに

対象位置に依存しないシフト不変特徴量は、ロバストな画像認識を行う上で有効な画像特徴量である。これまでに、HLAC 特徴[1]、SIFT 特徴[2]、HOG 特徴[3]といった様々な形の不変特徴量が提案されており、様々な応用システムに適用されてきた[4]–[6]。これらの多くは、NTSC 信号(30fps)等の人間の視覚特性に対応したビデオ信号規格を用いた画像システムを用いており、人間の目では見ることができない高速現象の実時間画像認識を行うためには限界がある。その意味で、FA 分野、バイオ分野を始めとした様々な分野における高速実時間画像認識の要求に十分な対応がされていないのが現状である。

そこで本研究では、シフト不変特徴量として HLAC 特徴を用いた高速画像認識システムを実現する。具体的には、HLAC 特徴を高速計算可能とする HLAC 回路モジュールを設計し、実際に高速ビジョンプラットフォーム H³ Vision 上にハードウェア実装することにより、1024×1024 画素画像に対し 1000fps での高速画像認識を実現する。またプロジェクトにより高速投影された画像に対する画像認識実験結果を示す。

2. 高次局所自己相関特徴

高次局所自己相関(HLAC)特徴[1]は、参照画素とその近傍画素に関する次の高次自己相関関数により計算される。

$$\int_D I(\mathbf{r})I(\mathbf{r} + \mathbf{a}_1)I(\mathbf{r} + \mathbf{a}_2) \cdots d\mathbf{r} \quad (1)$$

D は特徴抽出を行う画像領域、 $\mathbf{r}$ は参照画素位置、 $I(\mathbf{r})$ は $\mathbf{r}$ における輝度値、 $\mathbf{a}_n(n = 1, 2, \dots)$ は参照画素と近接画素の距離を表す。本研究では、近傍画素を3×3画素領域とし、0次、1次、2次のHLAC特徴に限定して議論を行う。

$$F_0 = \sum_D I(\mathbf{r}) \quad (2)$$

$$F_1 = \sum_D I(\mathbf{r})I(\mathbf{r} + \mathbf{a}_1) \quad (3)$$

$$F_2 = \sum_D I(\mathbf{r})I(\mathbf{r} + \mathbf{a}_1)I(\mathbf{r} + \mathbf{a}_2) \quad (4)$$

ここで計算するHLAC特徴は、0次: 1個、1次: 4個、2次: 20個の計25個である。参照画素を中心とした25個の局所パターンはFig. 1のようになる。

HLAC特徴は、近傍画素における輝度値同志を乗算した上で、画像全体に対する総和計算により求められるため、ハードウ

ェア回路実装を行う上で、回路リソースを多く必要とする乗算回数を抑えることが有効となる。

本研究では、例えば1次、2次のHLAC特徴の式(3)、(4)において、共通乗算項があることに着目する。2次HLAC特徴の乗算項は、1次HLAC特徴の乗算項を用いて、以下のように表わされる。

$$\varepsilon_1(\mathbf{a}_1) = I(\mathbf{r})I(\mathbf{r} + \mathbf{a}_1) \quad (5)$$

$$\varepsilon_2(\mathbf{a}_1, \mathbf{a}_2) = I(\mathbf{r})I(\mathbf{r} + \mathbf{a}_1)I(\mathbf{r} + \mathbf{a}_2) = \varepsilon_1(\mathbf{a}_1)I(\mathbf{r} + \mathbf{a}_2) \quad (6)$$

乗算結果の再利用を行わない場合、HLAC特徴を計算するために、全部で44回(1次: 4回、2次: 40回)の乗算が必要とされる。一方で、2次HLAC特徴の乗算項では、Fig. 2のように、1次HLAC特徴の乗算結果を17回分再利用可能である。また2次HLAC特徴のうち3個が、1次HLAC特徴の乗算結果を再利用できないが、この中で共通の乗算項を含む1組のHLAC特徴があるため、1回分の乗算結果を共有できる。その結果、乗算結果の再利用を行う場合、乗算回数は18回分抑えることができる。

Fig. 1: 25 local patterns for HLAC features

Fig. 2: Reuse of multiplied values in 1st order HLAC

3. システム実装

3.1 高速ビジョンプラットフォーム H³ Vision

次に前節で述べたHLAC計算アルゴリズムを高速ビジョンプラットフォーム H³ Vision[7]上に実装し、1024×1024 画素画像に対し1000fpsでの高速画像認識を実現する。H³ Visionは高速カメラヘッド、専用FPGAボード、パソコンからなる。

画像は高速カメラヘッドから16画素並列で専用FPGAボードへ転送される。専用FPGAボードは、高速通信・処理のためにDDR-SDRAMとFPGA(Xilinx Virtex II Pro. XC2VP100)間に160bitバスを用意し、1024×1024画素10bitグレイレベル画像に対し1000fpsでの処理を実現可能としている。これらの処理結果は、PCI-Xバスを介して、PC(CPU: Intel Xeon Dual Core 2.8GHz, 2GB memory)に転送される。パソコン上では任意の画像処理をソフトウェア実現可能である。

3.2 HLAC回路モジュール

本研究では、1024×1024画素といった画素レベルでの計算を必要とするHLAC特徴計算をハードウェア回路として実装し、最近傍法などの最終的な認識方法についてはソフトウェアで実装する。本節ではHLAC回路モジュールについて述べる。

HLAC回路モジュールは、1024×1024画素画像は64×64画素画像に分割され、全てのブロックに対して25個のHLAC特徴を1000fpsで計算する。Fig. 3, Fig. 4は、それぞれ、HLAC回路モジュールのブロックダイアグラム、制御信号のタイミングチャートを示す。入力画像は16画素並列で転送され、10bit階調1024×1024画素画像が、90MHzのクロックに従うX-address, Y-address信号により、Fig. 5(a)のように左上から右下へラスタ走査される。HLAC回路モジュールは、16画素並列入力された画像に対応した形で、並列動作を行う。

HLAC回路モジュールは16個の前処理サブモジュール、16個の乗算サブモジュール、25個の総和サブモジュールを持つ。

前処理サブモジュールでは、ノイズ除去等のために、入力画像から上位8bit階調値を選択する回路、及び3×1画素メディアフィルタからなり、16画素並列で動作する。

Fig. 3: Block Diagram of HLAC circuit module

Fig. 4: Timing Chart of HLAC circuit module

乗算サブモジュールでは、0次、1次、2次HLAC特徴における $I(r)$, $\varepsilon_1(a_1)$, $\varepsilon_2(a_1, a_2)$ を計算し、近傍画素の輝度値を乗算する。Fig. 6は乗算サブモジュールのブロックダイアグラムである。サブモジュールは、a) 0次、1次HLAC特徴に関する乗算を行う4つの乗算器(MUL1), b) MUL1の乗算結果を再利用し、2次HLAC特徴の乗算を行う22個の乗算器(MUL2)の2つの部分からなり、1サブモジュールあたり26個の乗算器から構成される。乗算サブモジュールは16並列で動作するため、HLAC回路モジュール全体では乗算器が416個実装される。なおこれらの全ての乗算結果が出力するために必要とされる遅延時間は11クロック(1クロック=11ns)である。

総和サブモジュールは、乗算サブモジュールで得られた乗算結果を加算し、64×64画素ブロック領域毎のHLAC特徴を計算する。Fig. 7は総和サブモジュールのブロックダイアグラムであり、a) 16画素分の乗算結果を加算する16入力加算器(ADD1), b) 画像のラスタ走査タイミングに同期し、ADD1により総和された画像を64×64画素ブロック単位で計算する逐次的加算器(ADD2)からなる。25個のHLAC特徴を並列に出力し、全HLAC特徴の出力に必要なとされる遅延時間は6クロックである。FPGAボード上のHLAC回路モジュールで64×64画素ブロック毎に計算された25次元のHLAC特徴は、Fig. 5(b)のようなブロック番号を付加した形でPCに転送される。

Fig. 5: Input image format and divided block regions

Fig. 6: Multiplication sub-module

Fig. 7: Summation sub-module

3.2 ハードウェア実装結果

HLAC 回路モジュールを H³ Vision ボード上の FPGA に回路実装した場合について、その回路リソース消費状況を Table 1 に示す。乗算回数に相当する 416 個の乗算器が並列実装され、FPGA リソース内に収まる設計が実現されていることがわかる。また 256 個の全ブロック領域からの 25 次元の HLAC 特徴について、PC に 1000fps で通信できることを確認している。

Table 1: Resource consumption of the FPGA on H³ Vision

Device Type	Xilinx Virtex Pro. XC2VP100
Slice	22,808/44,906 (51%)
Silce Flip Flop	38,321/88,192 (43%)
4 input LUT	16,405/88,192 (18%)
Bounded IOB	1,023/1,164 (87%)
Block RAM	137/444 (30%)
MULT18X18	416/444 (93%)
GCLK	7/16 (43%)
DCM	3/12 (25%)

4. 実験結果

4.1 実験環境

HLAC 特徴に基づく実時間画像認識を検証するために、プロジェクタにより高速投影された画像パターン認識実験を行った。Fig. 8 に実験環境を示す。プロジェクタには、ViALUX 社の ALP-3 コントローラボードの使用により秒間 8000 枚の XGA の 2 値画像を投影できる、TI 社の Discovery 1100 を用いた。

投影する画像パターンとしては、a) 0~9 の 10 枚の数字画像 (投影速度 500fps まで)、b) 10 枚の顔画像 (投影速度 250fps) を用いた。256 個の 64×64 画素領域で計算される HLAC 特徴は、1024×1024 画素の全画像に対する HLAC 特徴を求めるために、PC 上でソフトウェア加算される。

全画像に対する 25 次元の HLAC 特徴は最近傍法 [8] により画像パターン認識に用いた。ここで、クラス i ($i=1, \dots, 10$) のプロトタイプは、クラス i の学習データにおける 25 次元 HLAC 特徴の平均ベクトル μ_i とした。判別関数は、クラス i の学習データにおける標準偏差 σ_i により正規化された、以下の入力ベクトル $\mathbf{x}$ とプロトタイプ μ_i の距離関数を用いた。

$$d_i(\mathbf{x}) = \frac{\|\mathbf{x} - \mu_i\|}{\sigma_i}$$

全クラスの距離を計算した上で、その値が最小となるクラスを入力が属するクラスとした。なお全クラスにおける距離がある閾値以上である場合は、未知状態とした。

Fig. 8: Experimental environment

4.2 数字画像の認識実験

最初に 100fps 及び 500fps で 2 値の数字画像を高速投影した。0 から 9 の数字に対応した 10 枚の投影画像は Fig. 9 で示す画像を用いた。また H³ Vision のフレームレート、シャッター速度、空間解像度は、1000fps、1ms、1024×1024 画素に設定した。実時間認識実験を行う前に、予め、オフラインで取得された 1 クラスあたり 10 枚の教師画像に対する 25 次元 HLAC 特徴を教師データとして計算するものとし、計 100 枚の参照画像に基づく最近傍法により数字を認識することとした。なお未知状態は、全距離関数が 0.5 以上を出力する場合とした。

Fig. 10, Fig. 11 は、0 から 9 の順番に Fig. 9 で示した数字画像を繰り返し投影した場合について、認識された結果を示すものとなった。なお数字画像の投影速度はそれぞれ 100fps、500fps に設定した。Fig. 12 は、投影速度 500fps の場合について、正規化された 10 個の距離関数 $d_i(\mathbf{x})$ の 25ms 間にわたる時間変化を示す。

投影速度 100fps、500fps の場合ともに、全ての数字画像が対応したクラスに正しく認識され、判別されたクラスの距離関数が、他のクラスの距離関数に比べて小さくなることわかる。なおパターン切換時に、未知状態として認識される時間があるが、これはカメラとプロジェクタが非同期で動作していることが要因の一つと考えられる。

Fig. 9: Number images for experiment

Fig. 10: Number-image recognition using HLAC features (100 fps projection)

Fig. 11: Number-image recognition using HLAC features (500 fps projection)

Fig. 12: Distance from prototype samples for number images (500 fps projection)

Fig. 15: Distance from prototype samples for face images (250 fps projection)

4.3 顔画像の認識実験

次に 250fps でグレイレベルの顔画像を高速投影した。投影する画像としては、AT&T Laboratories Cambridge の顔データベース [9] から Fig. 13 の 10 枚の顔画像 (H1~H10) を選択した。なお 10 枚の顔画像は高速投影のため 8 階調に変換した。H³ Vision のフレームレート、シャッター速度、空間解像度は、500fps、2ms、1024×1024 画素に設定した。実時間認識実験を行う前に、オフラインで取得された 1 クラスあたり 100 枚の教師画像に対する 25 次元 HLAC 特徴を教師データとし、計 1000 枚の参照画像に基づく最近傍法により数字を認識する。未知状態の判定は数字画像の場合と同じとした。

Fig. 14 は、H1 から H10 の順に顔画像を繰り返し投影した場合の数字認識結果である。Fig. 15 は、正規化された 10 個の距離関数 $d_i(\mathbf{x})$ の 25ms 間にわたる時間変化を示す。全ての顔画像が対応したクラスに認識され、判別されたクラスの距離関数が、他のクラスの距離関数に比べて小さくなることからわかる。

これらの実験結果から、高速投影された画像パターンに対しても、1000fps で HLAC 特徴を高速ビジョンプラットフォーム上で計算することにより、グレイレベルに対する実時間画像パターン認識が実現されていることがわかる。

Fig. 13: Face images for experiment

Fig. 13: Face-image recognition using HLAC features (250 fps projection)

5. おわりに

本研究では、HLAC 特徴を用いて、1024×1024 画素画像に対する実時間画像パターン認識を 1000fps で実現した。HLAC 特徴を高速計算するための HLAC 回路モジュールを設計し、高速ビジョンプラットフォーム H³ Vision 上へハードウェア実装を行った。また高速プロジェクタにより投影された数字画像や顔画像に対する実時間画像パターン認識実験を行い、数百 fps レベルで高速に切り替わるグレイレベル画像に対しても、これらの画像パターンが正しく認識できることを確認した。

文 献

- [1] N. Otsu et al., A new scheme for practical, flexible, and intelligent vision systems, *Proc. IAPR Workshop on Computer Vision*, pp.431-435, 1988.
- [2] D. G. Lowe, "Object recognition from local scale-invariant features", *Proc. IEEE Int. Conf. on Computer Vision*, pp. 1150-1157, 1999.
- [3] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp.886-893, 2005.
- [4] M. Bicego, A. Lagorio, E. Grosso, and M. Tistarelli, "On the use of SIFT features for face authentication," *Proc. IEEE Int. Workshop on Biometrics*, pp.35, 2006.
- [5] T. Kobayashi, A. Hidaka, and T. Kurita, "Selection of Histograms of Oriented Gradients Features for Pedestrian Detection," *Proc. 14th Int. Conf. on Neural Information Processing*, pp.598-607, 2007.
- [6] T. Nanri, and N. Otsu, "Unsupervised Abnormality Detection in Video Surveillance," *Proc. IAPR Conf. on Machine Vision*, pp.574-577, 2004.
- [7] 石井, 山本, "超速ハイパービューマンビジョンとその応用," 電子情報通信学会誌, Vol.90, No.10, pp.838-841, 2007.
- [8] T.M. Cover and P.E. Hart, "Nearest neighbor pattern classification," *IEEE Trans. on Information Theory*, Vol.13, No.1, pp.21-27, 1967.
- [9] F.S. Samaria and A.C. Harter, "Parameterisation of a stochastic model for human face identification," *Proc. 2nd IEEE Workshop on Applications of Computer Vision*, pp.138-142, 1994.