2009－CVIM－166 (2)
2009／3／13

サポートベクターマシンを用いた Bag-of-Features における局所特徴の初期特徴選択

鈴木 浩二† 松川 徹† 栗田 多喜夫‡

†筑波大学大学院システム情報工学研究科 〒305-8577 茨城県つくば市天王台 1-1-1
‡産業技術総合研究所脳神経情報部門 〒305-8568 茨城県つくば市梅園 1-1-1 つくば中央第 2
E-mail: {ko-suzuki, t.matsukawa, takio-kurita}@aist.go.jp

あらまし 局所特徴を SVM により背景領域からの特徴と対象領域からの特徴に分類し利用する手法(初期特徴選択)を提案する．初期特徴選択の有効性を検証するために，UIUC Image Database を用いて SIFT 特徴を車領域からの SIFT 特徴と背景領域からの SIFT 特徴に分類する SVM を作成し，SIFT 特徴を用いた Bag-of-Features の特徴選択に適用する実験を行った．この実験の結果，Bag-of-Features において背景領域からの SIFT 特徴を SVM で選択的に取り除くことにより，クラスタ数が少ない場合において従来手法よりも識別率を向上させることできた．

キーワード Bag-of-Features，SVM，SIFT，初期特徴選択，一般物体認識

Bag-of-features car detection based on selected local features using Support Vector Machine

Koji SUZUKI† Tetsu MATSUKAWA† and Takio KURITA‡

† University of Tsukuba, Graduate School of Systems and Information Engineering
1-1-1 Tennoudai, Tsukuba, Ibaraki, 305-8577 Japan

‡ National Institute of Advanced Industrial Science and Technology
AIST Central 2, 1-1-1 Umezono, Tsukuba, Ibaraki, 305-8568 Japan
E-mail: {ko-suzuki, t.matsukawa, takio-kurita}@aist.go.jp

Abstract We propose a local feature selection method that classifies local features into background's features and target's features using SVM. We applied this feature selection method to "bag-of-features" method in generic object recognition problem. To verify the effectiveness of the proposed method, we conducted experiments using UIUC Image data base. Experimental results showed the proposed method outperformed the conventional Bag-of-Features representation with a fewer number of clusters.

Keyword Bag-of-Features, SVM, SIFT, preliminary feature selection, generic object recognition

1. はじめに

近年の PC の普及や処理能力の向上に伴い，画像認識技術は大きな進歩を遂げている．中でも，大量の画像データから目的の物体が写っている画像を認識する一般物体認識の技術は，Web 画像データマイニング，情報セキュリティ，ロボット視覚，デジタル家電など多くの分野において応用されてきており，盛んに研究が行われている[3,4,5,8,9,10].

こうした中で，Bag-of-Features と呼ばれる一般物体認識手法が G. Csurka らによって提案され[3]，近年多くの研究者が改良手法を提案している[8, 9, 10].

Bag-of-Features はテキストマイニングにおける Bag-of-Words 手法を画像に応用した手法であり，画像を画像から抽出される特徴量の集合として捉える．すなわち，ベクトル量子化した画像パッチのヒストグラムを用いた統計的パターン認識手法により画像認識を行う手法である．Bag-of-Features の認識に用いられるパターン認識手法には生成モデルと判別モデルが用いられている．L. Fei-Fei らは，特徴量に lowe らによって考案された SIFT 特徴量[6]を用い，クラスタリング手法の k-means 法と確率的トピック分析手法の LDA(Latent Dirichlet allocation)[2]を用いた手法により 13 種類の画像の識別を行い，64%の識別率を達成して

いる[4]. LDA が生成モデルである一方，判別モデルの代表手法として SVM(Support Vector Machine)と呼ばれる識別アルゴリズムがある．この手法は，現在知られているパターン認識手法の中でも，特に認識性能が優れている手法として知られており，Bag-of-Features で用いられる識別手法の多くがこの SVM を用いて行われている[5, 8, 10].

本稿では Bag-of-features 手法の改良手法として，認識アルゴリズムの初期段階で特徴選択を行う手法，**初期特徴選択**を提案する．従来の Bag-of-Features では，入力画像から求めた SIFT 特徴をすべて用いてヒストグラムを作成していたのに対して，提案手法では，入力画像から求めた SIFT 特徴をひとつずつ SVM に入力し，SVM で背景部分から抽出された SIFT 特徴と識別された SIFT 特徴を除外してヒストグラムを作成する．UIUC Image Database を用いた車両認識実験を行い，提案手法が，クラスタ数が少ない場合に識別性能が従来手法よりも向上することを示す．

2. Bag-of-Features

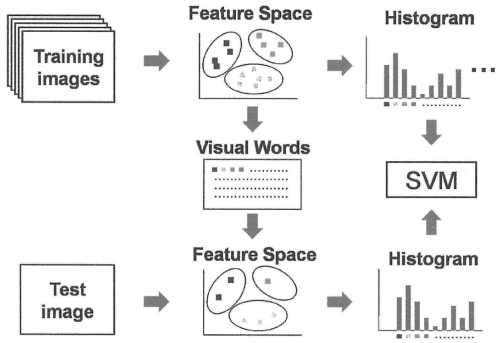


図 1. Bag-of-Features 手法概要

Bag-of-Features では，処理は学習フェーズと識別フェーズに分類される．そのため入力画像も図 1 のように学習用画像と識別用画像別々に用意する必要がある．

画像は局所特徴量の集合として捉えられるため，最初に画像から局所特徴量を抽出する必要がある．学習フェーズでは，複数枚の学習用画像から抽出された局所特徴量をまとめて特徴空間内でクラスタリングを行い，各クラスタの重心点を Visual Words と呼ばれる画像を表すための辞書として登録する．次に，各学習用画像一枚から特徴量のベクトル量子化を行い，Visual Words の出現頻度をカウントしたヒストグラムを作成する．そして，SVM に作成したヒストグラムを学習サンプルとして入力することにより，SVM の学習を行う．識別フェーズでは，同じように識別用画像から求められた特徴量を学習フェーズで作成した Visual Words を基にヒストグラムを作成し，SVM に識別サンプルとして入力することにより識別結果を得ることが出来る．

特徴量には SIFT 特徴を用いた手法が高い識別精度である場合が多いという実験結果もあり[7]，一般的にもこの SIFT 特徴がよく用いられるため，本実験では局所特徴量としてこの SIFT 特徴を使用して実験を行う．また，クラスタリング手法には k-means 法を用い，識別の際の SVM には linear SVM を用いて実験を行う．

3. 提案手法概要

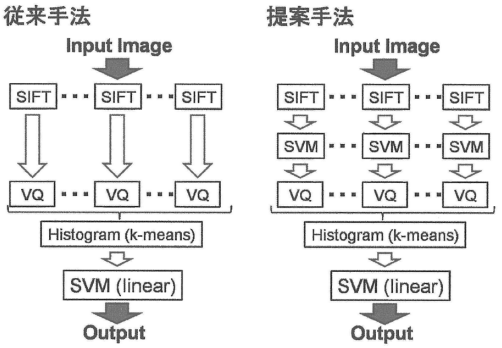


図 2. Bag-of-Features の従来手法と提案手法の比較

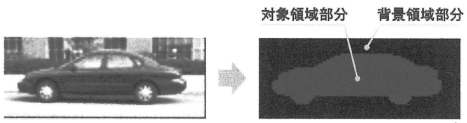


図 3. 初期特徴選択基準

上記のように Bag-of-Features は，求めた特徴量の位置情報を無視して画像を特徴の集合として捉えるため，画像内の背景領域や物体領域といった位置情報を無視しているといえる．S. Lazebnik らは Bag-of-Features の改良手法として局所特徴の位置を考慮して，ヒストグラムを分割する Spatial Pyramid Matching 手法を提案し，通常の Bag-of-Features より高い精度で識別が可能になることを示している[5]．このことから Bag-of-Features に位置情報を用いることは有効な手段であると考えられる．

本稿では，図 2 のように Bag-of-Features における局所特徴量(SIFT 特徴)を，SVM を用いて初期段階で領域情報を基に選択する手法を提案する．従来の Bag-of-Features では，入力画像から求めた SIFT 特徴をすべて用いてヒストグラムを作成していたのに対して，提案手法では，入力画像から求めた SIFT 特徴をひとつずつ SVM に入力し，SVM で背景部分から抽出された SIFT 特徴と識別された SIFT 特徴を除外してヒストグラムを作成する．この SVM による SIFT 特徴の特徴選択は，学習フェーズ，識別フェーズ両方の段階で行

われ，入力画像のすべてのクラスに行われる．また，この SVM の識別基準には，図 3 のようなマスク画像を用いて，求められた SIFT 特徴が対象物体領域部分から抽出されたものか，背景領域部分から抽出されたものか，という 2 クラスの識別基準を用いる．そして SVM により対象物体領域部分として識別された SIFT 特徴のみを用いてヒストグラムの作成を行う．

従来の Bag-of-Features の特徴選択では，ベクトル量子化された後の特徴量に対して行われていた[9,10]のに対して，本研究の提案手法ではベクトル量子化する以前の初期段階の特徴量に対して特徴選択を行う．本稿ではそのような，クラスタリング等の教師無し学習による特徴変換以前の特徴選択を“**初期特徴選択**”と呼び，SVM を用いて実現する．この初期特徴選択が従来手法と違う点は，対象物体の領域というトップダウンな情報を用いて，低次の特徴量に対する特徴選択を行う，というところにある．

4. 実験

4.1. 実験条件

Bag-of-Features の手法としては，局所特徴量の特徴点検出に DoG 画像によるキーポイント検出法を用い，特徴記述には SIFT 記述子を用いた．また，ベクトル量子化にはクラスタリング手法として k-means 法を用い，ヒストグラムは Visual Words の頻度をカウントした値を正規化せずに入力値に用いた．画像を分類するための識別器には linear SVM を用いた．また，SIFT 特徴を識別する際の SVM には linear SVM と kernel SVM を用い，kernel SVM のカーネルには Gaussian kernel を使用した．SVM のハイパーパラメータは 5 分割交差確認法による格子点検索でパラメータ探索を行い決定した．SIFT 特徴の算出には R. Hess 氏[11]が公開しているソースコードを使用し，SVM には C. -J. Lin 氏らよって作成されたライブラリ LIBSVM[12]を使用した．

SVM による初期特徴選択で SVM に入力する特徴量は，以下の 3 種類を用いた．

- **SIFT 特徴 (*S_lin, S_ker*)**
  128 次元のエッジベースのヒストグラム特徴量をそのまま SVM の入力値として使用した．また SVM のカーネルには linear SVM(*S_lin*)と kernel SVM(*S_ker*)の二つを使用した．

- **SIFT 特徴+輝度値 (*S+lum*)**
  SIFT 特徴に加えて図 4 のように SIFT 特徴と同じ領域から 12x12 領域のピクセルの輝度値の平均値を算出し，144 次元の輝度値の特徴量と 128 次元の SIFT 特徴を合わせて全 272 次元の特徴量を作成し入力値として使用した．SVM には kernel SVM のみを使用した．

- **SIFT 特徴+輝度値+SIFT 特徴のスケール情報+SIFT 特徴の位置情報 (*S+lum+scl+pos*)**
  上記 SIFT 特徴と輝度値に加え，SIFT 特徴を算出する際の DoG 画像によるスケール情報(s)とキーポイントの位置情報(x,y)の計 3 次元の特徴量を追加し，全 275 次元の特徴量として入力値に使用した．

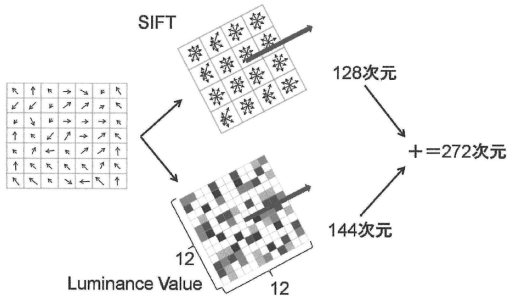


図 4. SIFT 特徴＋輝度値情報

本研究では，SVM による初期特徴選択を行わない場合を従来手法(*conventional*)として比較実験を行った．

4.2. データセット

表 1. 入力画像とマスク画像の例

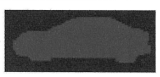
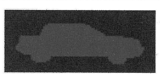
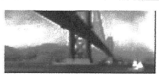
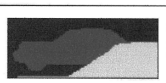
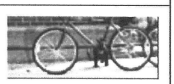
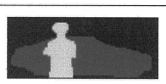
Positive Sample		Negative Sample
src	mask	src
		
		
		
		

表 1 に本研究で用いた入力画像と作成したマスク画像の例を示す．

入力画像は UIUC Image Database for Car Detection[13]の 1050 training images(550 car and 500 non-car images)を用いた．この画像データベースは 550 枚の車画像(Positive Sample)と 500 枚の非車画像(Negative Sample)から成り，画像のサイズは全て

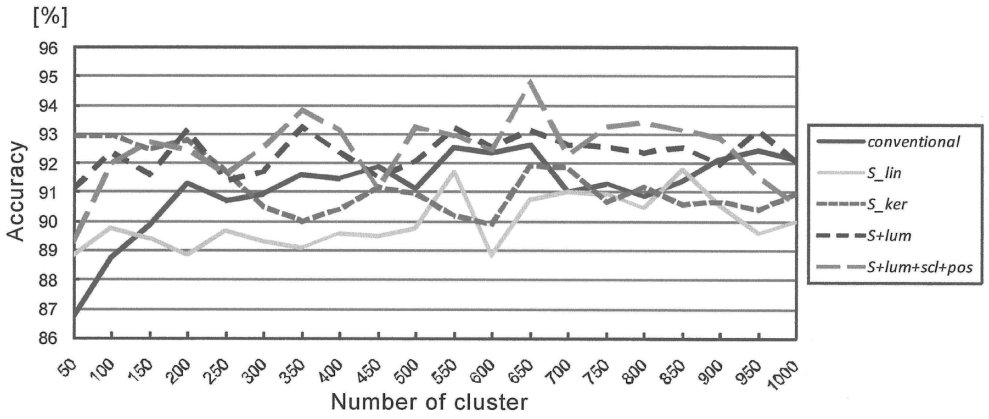


図 5. 画像識別結果

100x40 (pixels)である．また，写っている車の大きさは全てスケーリングされ揃えられている．実験ではこれらの画像を，車画像か非車画像かの 1-vs-rest の問題として Bag-of-Features の認識課題を設定し，3 分割交差確認法で評価した．また，初期特徴選択のクラスラベル作成のために表 1 のようなマスク画像を全ての車画像について作成し，自動的にクラスラベル付けを行った．また，入力画像の中には図のようにオクルージョンがある画像も存在する．それらオクルージョンもマスク画像に含め，オクルージョン領域部分は背景領域部分としてクラスラベルを作成し実験を行った．

4.3. 実験結果

4.3.1. SIFT 特徴の初期特徴選択結果

表 2. SIFT 特徴の識別結果(Accuracy)

	S_lin	*S_ker*	*S+lum*	*S+lum+ scl+pos*
1	63.64%	80.72%	88.69%	**91.08%**
2	62.94%	80.92%	88.64%	**91.44%**
3	64.48%	80.72%	88.30%	**91.18%**
average	63.69%	80.79%	88.54%	**91.23%**

入力画像一枚あたりから抽出された SIFT 特徴の数は平均して約 41 個であった．

表 2 に SVM で車画像の SIFT 特徴を車領域部分から抽出された SIFT 特徴と背景領域部分から抽出された SIFT 特徴に選択した識別結果を示す．縦にそれぞれ 3 分割交差確認法における 3 回の結果とそれらの平均値を示している．

3 回の平均値を見ると SIFT 特徴をそのまま Linear SVM で識別した識別性能が一番低く 63.69%となっている．また，SIFT 特徴のみを用いて識別するよりも，輝度値や位置といった特徴量を加えた方が識別性能が高くなり，SIFT 特徴に輝度値，スケール情報，位置情報を加えた結果(*S+lum+scl+pos*)が平均識別率 91.23%と一番高い識別精度になっていることが分かる．SIFT 特徴に輝度値を加えた特徴量は，SIFT 特徴と同じ領域から算出しているため，スケールと回転に不変な特徴量になっている，しかし，スケール情報と位置情報はそのような SIFT 特徴の特性にはよらない特徴である．本研究で用いた画像セットはスケーリングされており，車が写っている領域も大体揃っていることから，スケール情報と位置情報を加えた特徴量で，識別率が上がっていると考えられる．

4.3.2. 画像の識別結果

図 5 に各手法で画像を識別した結果を示す．

本実験では画像一枚あたりから抽出される SIFT 特徴が約 41 個であることから，クラスタ数は 1000 までで十分であると考えクラスタ数の範囲を設定した．

図 5 を見ると，従来手法ではクラスタ数が小さいところでは識別精度が低い値になっているが，クラスタ数が大きくなるにつれ，識別精度が上がっていき，クラスタ数が 200 あたりでおよそ一定の値になっていることが分かる．また，4 つの提案手法と従来手法を比較すると，いずれの場合でもクラスタ数が小さいところでは従来手法よりも提案手法の精度が勝っていることが分かる．また，クラスタ数が大きいところでは SIFT 特徴を linear SVM で識別した手法(*S_lin*)以外の 4 つの手法の精度に大きな差はないように見える．SIFT 特徴を linear SVM で識別した手法(*S_lin*)は，クラスタ数全体での識別性能を見ると従来手法よりも識別精度が若干低いところが多いように見える．これは SIFT 特徴の識別性能が低く，車領域部分からの SIFT 特徴も多く削除してしまったためだと考えられる．また，

SIFT特徴+輝度値(*S+lum*)とSIFT特徴+輝度値＋スケール＋位置(*S+lum+scl+pos*)を初期特徴選択に用いた場合，クラスタ数を大きくしても従来手法よりも若干性能が向上している．画像識別性能は，*S+lum+scl+pos*, *S+lum*, *S_ker*, *S_lin* の順に高く，これは最終的な画像識別性能に初期特徴選択の性能が影響していることを示している．

4.4. Linear SVMによるヒストグラムの重み付け結果

次に，より詳細に識別結果を調べるために，従来手法(*conventional*)とSIFT特徴をkernel SVMで識別する手法(*S_ker*)とSIFT特徴をlinear_SVMで識別する手法(*S_lin*)で，画像を識別する際のlinear SVMで学習されたヒストグラムの各ビンの重みを算出し考察する．

図6に一枚の車画像から求められたヒストグラムのlinear SVMによる重み付け結果のグラフを示す．また，図7に一枚の非車画像から求められたヒストグラムのlinear SVMによる重み付け結果のグラフを示す．これらのグラフは画像を識別する際のlinear SVMが学習した重みの大きい順でソートされている．ソートの順番は左側が車領域部分から抽出されたSIFT特徴として大きな重みを付けられたもの，右側が背景領域部分から抽出されたSIFT特徴として大きな重みを付けられたものという順になっている．また，各ヒストグラムの値はマスク画像によって付けられた，車領域部分のSIFT特徴と背景領域部分のSIFT特徴のクラスラベルで色分けされている．

図を見ると，図6では，クラスタ数が50のときに，従来手法(a1)では同じクラスタ内に車領域部分のSIFT特徴と背景領域部分のSIFT特徴が混ざってしまい，上手く左右に重み付けできないクラスタが出来てしまっていることが分かる．これに対し，提案手法のkernel SVMを用いた手法(a2)では，背景領域部分のSIFT特徴をよく削減できているために，同じクラスタ内に二つのクラスラベルのSIFT特徴が混ざることが少なくなって，左側方向に車領域部分のSIFT特徴が集まり，上手く重み付けすることが出来ていることが分かる．また，提案手法のlinear SVMを用いた手法(a3)では，背景領域部分のSIFT特徴を十分削減出来ていない上に，車領域部分のSIFT特徴も間違えて削減してしまっているためにあまり上手く重み付けが出来ていない状態であることが分かる．

しかし，クラスタ数が300になると，各クラスタ内の値は同じクラスタに重複することなく分類されるようになるため，従来手法(b1)，提案手法のkernel SVMを用いた手法(b2)，linear SVMを用いた手法(b3)，それぞれで上手く重み付けが出来ていることが分かる．

さらに，非車画像から求められたヒストグラムのlinear SVMによる重み付け結果のグラフを見ても，同じように，クラスタ数が50の時には従来手法(c1)では上手く右側にSIFT特徴を集められていないのに対して，提案手法のkernel SVMを用いた手法(c2)とlinear SVMを用いた手法(c3)ではより右側にSIFT特徴を集めることが出来ており，上手く重み付け出来ていることが分かる．そして，クラスタ数が300のときを見ると，従来手法(d1)も提案手法(d2)(d3)と同じように右側にSIFT特徴が集まるようになり，重み付けが上手くいっていることが分かる．

以上のことから，クラスタ数が少ない場合に，提案手法が識別率が高い理由は，同クラスタ内に存在する背景領域部分のSIFT特徴を削減することにより，クラスタの重み付けを適正に行えるようになるため，クラスタ数が少ない場合に，提案手法の識別率が向上したと考えられる．

5. 結論

認識アルゴリズムの初期段階で特徴選択を行う手法，**初期特徴選択**を提案した．本稿ではBag-of-FeaturesにSVMを用いた初期特徴選択を適用することによりクラスタ数が少ない場合に識別精度が向上することを示した．クラスタ数が少ない場合に識別精度が高いことは，後段で行うSVMに入力する特徴次元の数を少なくすることができ，後段の処理の高速化，メモリ削減につながる手法であるといえる．

実験の結果は，用いた画像セットに依存する結果である可能性がある．そのため，その他の画像セットに提案手法を適用し，本研究の結果が一般的に成り立つか検証する必要がある．また，提案手法はSURF[1]やGLOH[7]といったSIFT以外の局所特徴量への応用が可能であると考えられるが，それらの手法への応用は今後の課題である．

文　献

[1] H. Bay, T. Tuytelaars, and L. V. Gool, “SURF: Speeded Up Robust Features,” *ECCV*, pp. 404-417, May 2006.

[2] D. Blei, A. Ng, and M. Jordan. “Latent dirichlet allocation,” *Journal of Machine Learning Research*, 3:993-1022, 2003.

[3] G. Csurka, C. Bray, C. Dance, and L. Fan, “Visual categorization with bags of keypoints,” *Proc. of ECCV Workshop on Statistical Learning in Computer Vision*, pp. 1-22, 2004.

[4] L. Fei-Fei, and P. Perona. “A Bayesian hierarchical model for learning natural scene categories,” *Proc. of IEEE Computer Vision and Pattern Recognition*, pp. 524-531, 2005.

[5] S. Lazebnik, C. Schmid, and J. Ponce, “Beyond bags of features, Spatial pyramid matching for recognizing natural scene categories,” *Proc. of IEEE Computer Vision and Pattern Recognition*, pp.

2169-2178, 2006.

[6] D. G. Lowe, “Object recognition from local scale-invariant features,” *Proc. of IEEE International Conference on Computer Vision*, pp. 1150-1157, 1999.

[7] K. Mikolajczyk, and C. Schmid. “A performance evaluation of local descriptors,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 27, No. 10, pp. 1615-1630, 2005.

[8] F. Perronnin, C. Dance, G. Csurka, and M. Bressan: “Adapted vocabularies for generic visual categorization,” *Proc. of European Conference on Computer Vision*, pp. IV:464-475, 2006.

[9] R. Sukthankar, L. Yang, R. Jin and F. Jurie, “Unifying discriminative visual codebook generation with classifier training for object category recognition,” *Proc. of IEEE Computer Vision and Pattern Recognition*, 2008.

[10] J. Yang, Y. G. Jiang, A. Hauptmann, and C.W. Ngo, “*Evaluating bag-of-visual-word* representation in scene classification,” *MIR’07 ACMMM*, 2007.

[11] http://www.csie.ntu.edu.tw/~cjlin/libsvm/

[12] http://web.engr.oregonstate.edu/~hess/

[13] http://l2r.cs.uiuc.edu/~cogcomp/Data/Car/

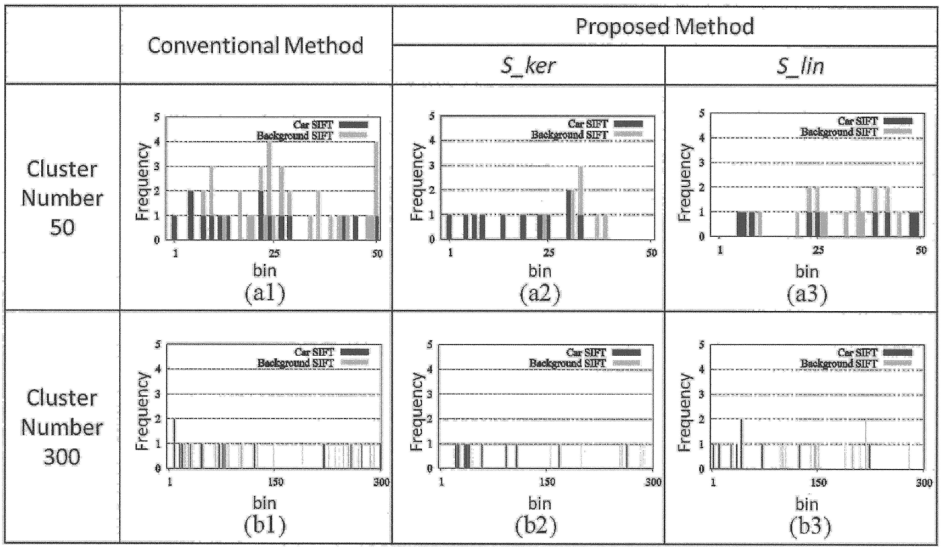


図 6. 車画像から求めたヒストグラム例

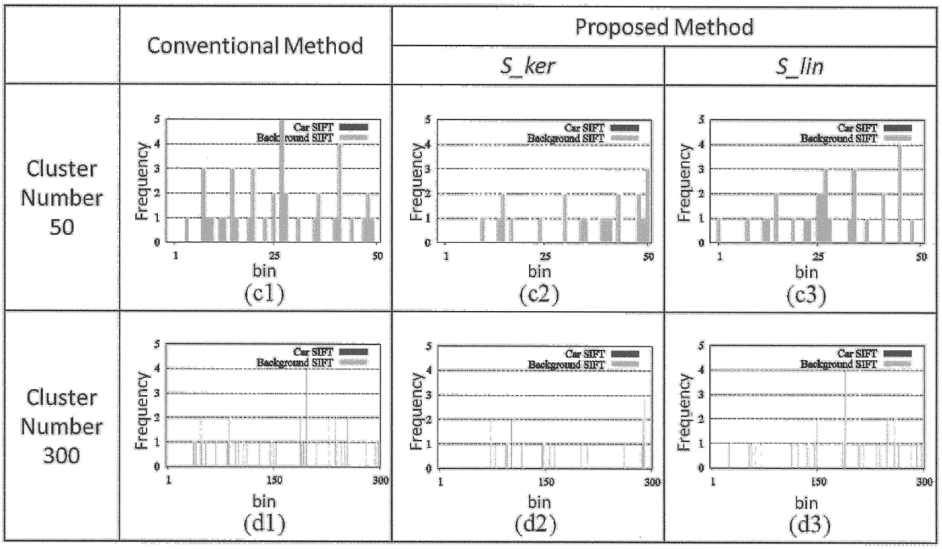


図 7. 非車画像から求めたヒストグラム例