

生成型学習法を用いた姿勢変化に頑健な歩行者検出の検討

吉田 英史[†] 出口 大輔[†] 井手 一郎[†] 村瀬 洋[†]

後藤 邦博^{††} 木村 好克^{††} 内藤 貴志^{††}

[†] 名古屋大学 〒464-8603 愛知県名古屋市千種区不老町

^{††} (株)豊田中央研究所 〒480-1192 愛知県愛知郡長久手町大字長湫字横道 41-1

E-mail: †yoshidah@murase.m.is.nagoya-u.ac.jp, ††{ddeguchi,ide,murase}@is.nagoya-u.ac.jp

あらまし 近年、車載カメラで撮影された画像から歩行者を検出する研究が注目されている。しかし、姿勢変化が大きく、かつ様々な背景に存在する歩行者を検出することは容易ではない。本研究では、車載カメラ画像からの歩行者検出において、多様な姿勢変化に対応するための手法を提案する。画像中の歩行者の見えを特徴とした従来の歩行者検出手法では、歩行者の姿勢変化や歩行者を取り巻く環境の変化に対応するために、学習用に事前に大量の歩行者画像を人手で収集する必要がある。提案手法ではこの問題に対して、少数の歩行者画像をいくつかの姿勢クラスに分類した後に、姿勢クラスごとに多様な歩行者画像を生成し、さらにこの姿勢クラスをテンプレートとしたマルチテンプレート型の識別器を構築することで解決を図る。実験の結果、従来手法に比べて提案手法の検出精度は大きく向上し、その有効性を確認した。

キーワード 歩行者検出, 生成型学習法, HOG, SVM

A study on a method for stable pedestrian detection against pose changes with generative learning

Hidefumi YOSHIDA[†], Daisuke DEGUCHI[†], Ichiro IDE[†], Hiroshi MURASE[†],

Kunihiro GOTO^{††}, Yoshikatsu KIMURA^{††}, and Takashi NAITO^{††}

[†] Nagoya University Furo-cho, Chikusa-ku, Nagoya, Aichi 464-8601 Japan

^{††} Toyota Central Research & Development Laboratories, Inc. Nagakute, Aichi, 480-1192, Japan

E-mail: †yoshidah@murase.m.is.nagoya-u.ac.jp, ††{ddeguchi,ide,murase}@is.nagoya-u.ac.jp

Abstract Recently, pedestrian detection from in-vehicle camera images is being focused. However, it is difficult to detect pedestrians due to the variety of their poses and backgrounds. To tackle this problem, we propose a method to detect various pedestrians from in-vehicle camera images. To deal with changes of pedestrians' pose and environment, most existing methods making use of their appearance require to prepare a lot of pedestrian images manually. The proposed method classifies a small number of pedestrian images into several pose classes and then generates various pedestrian images from each pose class. Finally, the proposed method constructs a classifier based on multiple templates from each pedestrian pose. Experimental results showed that the detection accuracy of the method outperformed existing methods, and we confirmed its effectiveness.

[Note] This document is an informal handout distributed at an IEICE TC-PRMU workshop.

Key words Pedestrian detection, Generative learning, HOG, SVM

1. ま え が き

近年、自動車運転支援技術の1つとして、車載カメラによって撮影された画像から歩行者を検出する研究が盛んに行われている。車載カメラ画像を用いた歩行者検出の多くは、歩行者と

その周囲のエッジや濃淡といった見えの情報を使用する。しかしながら、画像中の歩行者の姿勢が大きく変化することや、複雑で多様な背景が存在すること、照明条件の違いにより明るさが変化することなど、歩行者を取り巻く環境の変化が歩行者検出を難しくしている。このような見えの変化の中でも、歩行者

の姿勢変化は画像中で検出対象の形状変化として現れるため、検出精度に大きく影響する。そこで、歩行者を取り巻く環境や歩行者の姿勢変化を考慮した、高精度な歩行者検出手法が求められている。

Dalal らは、HOG 特徴量と SVM を用いた歩行者検出手法を提案している [1]。HOG 特徴は画像の局所領域の輝度勾配方向ヒストグラムであり、照明変動や影の影響を受けにくく、局所的な幾何学変化に頑健である。また、SVM は高い汎化性能を持つ識別器である。この手法では、歩行者の姿勢変化に対処するために様々な姿勢、環境下の歩行者画像を多数収集する必要がある。しかしながら、このような歩行者画像を人手で収集することは多くの時間やコストを要する。

そこで本研究では、学習に使用する歩行者画像を少量の歩行者画像から体系的に生成することを考える。この考え方は、“生成型学習法”として様々な対象に適用されている [2]～[7]。生成型学習法は、実際の環境で起こりうる様々な見えの変動をモデル化し、このモデルを用いて少量の原画像から多様な変動を含む大量の生成画像を体系的に生成する。これにより、実際に対象となる画像を収集する場合よりも多様な変動を含む画像を学習に使用できるようになり、学習用画像の収集コストを大幅に削減することができる。従来、生成型学習法は文字や一般物体 [2]、道路標識 [3]、路面表示 [4] を対象として研究されており、各対象に適した生成モデルが提案されている。この生成型学習法の考え方は歩行者検出にも利用されている。しかしながら、従来の生成型学習法を用いた歩行者検出 [5]～[7] では歩行者の姿勢変化や周辺環境を十分に考慮していないため、さらなる精度向上が求められている。

本研究では歩行姿勢形状をいくつかの姿勢クラスに分類した後、各姿勢クラス毎に歩行者画像を生成し、識別器もその姿勢クラス毎に構築することで高精度な歩行者検出の実現を目指す。

以降、2 節で提案手法の流れについて述べ、学習用歩行者画像の生成手順、マルチテンプレート型の識別器の構築方法について詳しく説明する。3 節では実際に車載カメラにより撮影された画像を用いた比較実験について述べ、考察を加える。最後に 4 節で本報告をまとめる。

2. 提案手法

図 1 に提案手法の流れを示す。提案する歩行者検出手法は、生成段階、学習段階、検出段階の 3 つからなる。

生成段階では、歩行者画像を歩行姿勢形状に基づくクラスタリングにより各姿勢クラスに分類し、姿勢クラス毎に大量の歩行者画像を生成する。歩行者画像の生成は、形状の生成、テクスチャの生成、背景の合成の順に行なう。学習段階では、歩行姿勢毎に生成した歩行者画像に非歩行者画像を加え、マルチテンプレート型の識別器を学習・構築する。検出段階では、学習段階で構築した識別器を用いて車載カメラ画像中から歩行者を検出する。

2.1 歩行者画像の生成

本節では、少量の歩行者画像から多数の学習用の歩行者画像を生成する生成処理について述べる。生成の種となる少量の歩

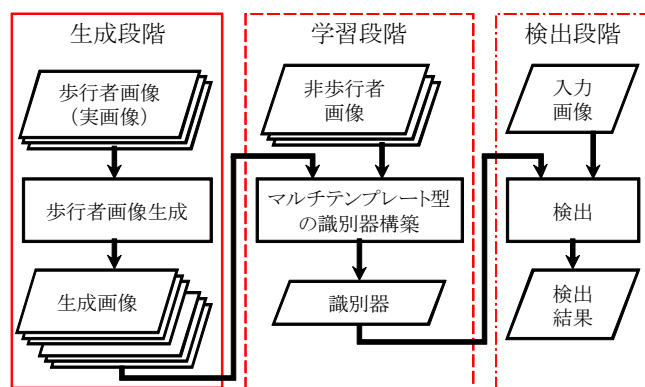


図 1 提案手法の流れ

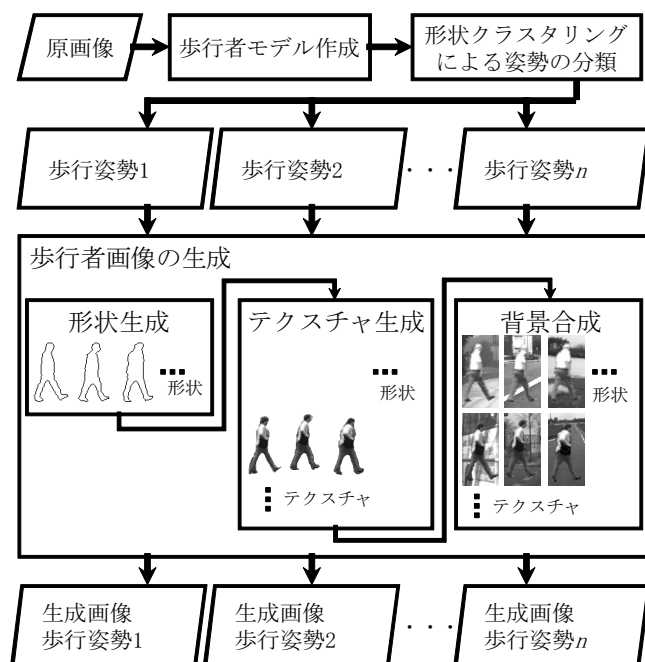


図 2 歩行者画像生成の流れ

行者画像を本報告では原画像と呼び、これらは人手によって収集する。一方で、生成処理によって得られる新たな歩行者画像を生成歩行者画像と呼ぶ。図 2 に生成処理の流れを示す。提案手法ではまず、歩行者画像を歩行者の形状に基づいてクラスタリングすることで姿勢クラスに分類し、統計的に現れやすい歩行姿勢形状を得る。そして、得られた歩行姿勢を持つ歩行者画像を基準とし、学習用の歩行者画像を生成する。生成処理は、歩行者の形状生成、テクスチャ生成、背景合成の順に行う。以降、各処理について詳細に述べる。

2.1.1 歩行者モデルの作成

まずはじめに、原画像から歩行姿勢形状を表す歩行者モデルを作成する。歩行姿勢形状は、図 3 に示すように原画像中の歩行者の輪郭上に等間隔にプロットした N 個の制御点の座標ベクトル $V = (x_1, y_1, \dots, x_N, y_N)^t$ により表す。以降これを制御点ベクトルと呼ぶ。なお、全ての原画像で制御点ベクトルの次元は等しいものとする。

2.1.2 形状クラスタリングによる歩行姿勢の分類

クラスタリング処理は、Gavrila らの手法 [8] を用いて行う。

原画像	
生成画像	

図 7 同じ歩行姿勢形状に異なるテクスチャを生成し合成した例

図 8 背景切り出し領域例

2.1.5 背景合成

最後に、歩行姿勢形状とテクスチャを生成した歩行者画像に多様な背景を貼りつける処理を行う。新たな背景は、歩行者の写っていない車載カメラ画像から収集する。まず、歩行者の写っていない画像に対してスケール変化を加え、歩行者画像と同じ大きさに切り出す。ここで、スケールに応じて切り出す領域を制限することで、空やアスファルトのみの画像を除外する。切り出し領域例を図 8 に示す。図 8 は、歩行者画像が図中右画像の大きさのとき、左画像の枠で囲まれた領域を $\Delta x_1, \Delta y_1$ で与え、その領域から背景を切り出すことを示す。

次に、生成した歩行者画像から歩行者領域のみを抽出し、 x 軸方向と y 軸方向にそれぞれ $\Delta x_2, \Delta y_2$ の位置ずれを与えて背景に人物領域のテクスチャを貼り付ける。ここで、人物領域のエッジの画素と背景の重なる画素をアルファブレンディングすることで、背景と合成する。

2.2 マルチテンプレート型の識別器

姿勢クラス毎に生成した歩行者画像に、非歩行者画像を加えてマルチテンプレート型の識別器を構築する。識別器の構成を図 9 に示す。

まず、姿勢クラス C^k に属する生成歩行者画像と非歩行者画像から特徴抽出し、画像特徴を用いて 2 クラス識別器 $f^k(IN)$ を構築する。ここで、 IN は識別対象画像とする。この識別器は姿勢クラスそれぞれに対して構築し、その総数は K である。次に、構築された 2 クラス識別器を用いてマルチテンプレート型の識別器 $F(IN)$ を次のように構成する。

$$F(IN) = \max\{f^1(IN), f^2(IN), \dots, f^K(IN)\} \quad (5)$$

$$\begin{cases} F(IN) \geq \epsilon \Rightarrow \text{歩行者} \\ \text{other} \Rightarrow \text{非歩行者} \end{cases}$$

ここで構築するマルチテンプレート型の識別器の特徴として、ある 2 クラス識別器が非歩行者であると識別した歩行者画像であっても、他の 2 クラス識別器では歩行者と識別することがで

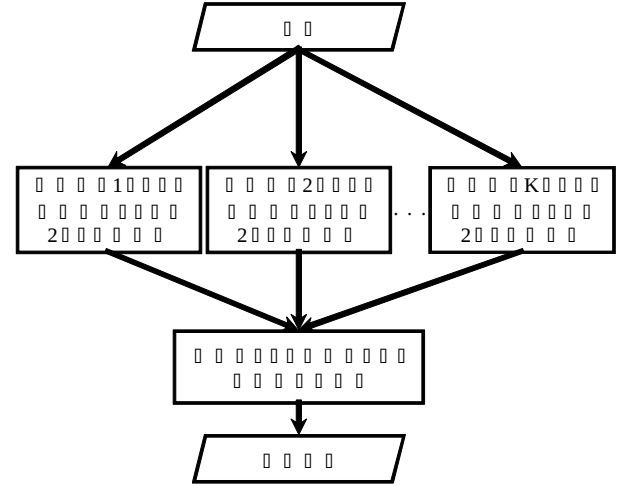


図 9 マルチテンプレート型の識別器



図 10 歩行者画像



図 11 非歩行者画像

きる。このような構造にすることで、それぞれの 2 クラス識別器が歩行者の複雑な姿勢変化に対応できるようにする。また、傘をさしている歩行者やしゃがんでいる歩行者などの新たな歩行姿勢の歩行者画像から学習した識別器を構成に加えることにより、容易に既存のマルチテンプレート型の識別器を拡張可能という特徴を有する。

3. 実験

本節では、提案手法の有効性を確認するために行った評価実験について述べる。

3.1 データセット

車載カメラから撮影された評価用画像系列や、識別器の構築に必要な学習用歩行者画像及び非歩行者画像は、Daimler 社により提供されているグレースケール画像のベンチマークデータセット [10] を利用した。

生成の種となる原画像には、図 10 のような歩行者画像 200 枚をランダムに収集して使用した。また非歩行者画像には、歩行者の写っていない車載カメラ画像にスケールを加え、歩行者画像と同じ大きさに切り出した図 11 のような画像を予備実験により 13,802 枚収集して使用した。使用した非歩行者画像の一部を図 11 に示す。

評価用画像は、ベンチマークデータセットに含まれる車載カメラ画像系列から収集した。このとき、80 pixels から 480 pixels の高さの歩行者を 1 人以上含む区間のみを評価の対象とした。

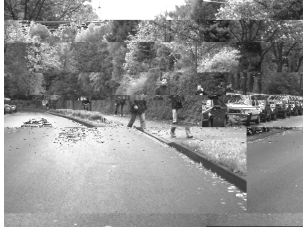


図 12 評価用画像の例



図 13 生成歩行者画像の一部

表 1 生成パラメータ 1

処理	パラメータ	値
歩行姿勢によるクラスタリング	N	100
	θ	380
	M	3
形状生成	n	2
	U	$45/S^k$
テクスチャ生成	m	20
	n'	2
	U'	$(m - M^k)/S^k$
背景合成	Δx_1	10 pixels
	Δy_1	± 5 pixels
	Δx_2	± 2 pixels
	Δy_2	± 4 pixels
	$\alpha(1)$	0.5
	$\alpha(2)$	0.8

この条件を満たす画像系列のうち、歩行者延べ人数 458 人を含む計 400 枚の画像を評価用画像として使用した。評価用画像の例を図 12 に示す。

3.2 学習用歩行者画像の生成

収集した原画像 200 枚を使用し、提案手法により様々な変化を加えた歩行者画像 9,867 枚を生成した。ここで歩行者の歩行姿勢形状は、10 の姿勢クラスに分類された。生成歩行者画像の一部を図 13 に示す。生成時に行った各生成処理については、表 1 に示すパラメータを用いた。歩行者モデルの作成における制御点のプロットは人手で行った。

表 2 手法概要

手法	生成	識別器構成	学習/生成型学習に用いた入力原画像枚数	識別器の構築に用いた画像枚数	
				歩行者	非歩行者
従来	なし	2 クラス	200	200	13,802
提案 1	あり	2 クラス	200	9,867	13,802
提案 2	あり	マルチ テンプレート型	200	9,867	13,802

3.3 識別器構築

生成した歩行者画像や非歩行者画像から抽出する画像特徴として HOG 特徴を使用し、学習に線形 SVM を用いた。SVM は、台湾国立大学の Lin らによって作られたライブラリパッケージである LibSVM [11] を利用した。

本実験では検出に要する計算量削減のため、予め疎な HOG 特徴を用いて 2 クラス識別器を原画像 200 枚と非歩行者画像 13,802 枚から別途構築した。この識別器によって歩行者であると識別された画像のみを、密な HOG 特徴を用いて構築した各手法の識別器に入力として与えた。

3.4 比較手法及び実験の概要

本実験で比較に用いた手法を表 2 にまとめる。提案手法 1 は本稿で述べた方法により原画像 200 枚から学習用の歩行者画像を生成し、得られた歩行者画像 9,867 枚と非歩行者画像 13,802 枚を用いて 2 クラス識別器を構築する手法である。提案手法 2 は、原画像 200 枚から生成によって得られた歩行者画像 9,867 枚と非歩行者画像 13,802 枚を用いて姿勢クラスごとにテンプレートとなる識別器を構築して、マルチテンプレート型の識別器を構築する手法である。従来手法は生成を行わず、原画像 200 枚と非歩行者画像 13,802 枚を用いて 2 クラス識別器を構成する手法である。これらの手法を用いて、車載カメラで撮影された画像系列に対し歩行者検出を行った。検出結果と人手で歩行者を囲った正解データとを比較して検出精度を調べた。

評価の指標には、ROC カーブを使用した。縦軸に検出率を、横軸に False Positive Per Frame (FPPW: フレームあたりの誤検出数) の対数値で示すことにする。式 (6) で示したパラメータ ϵ を変化させて描画した。検出によって描画された曲線が左上にシフトするほど検出精度が高いことを表す。

3.5 実験結果と考察

比較した 3 つの手法による歩行者検出の精度を表す ROC カーブを図 14 に示す。また、提案手法と比較手法 2 における実際の検出結果画像を図 15, 16 に示す。

図 14 から、提案手法 1 及び提案手法 2 は従来手法に比べて検出精度が大きく向上していることがわかる。これは学習用の歩行者画像を生成したことによる効果である。よって、歩行姿勢形状や歩行者のテクスチャ、背景をコントロールすることで様々な姿勢、テクスチャ、背景を持つ画像を用いた学習が可能となり、従来手法に比べて姿勢変化や背景変動に頑健な識別器を構築できたと言える。また、提案手法 2 の検出精度が提案手法 1 よりも高いことから、生成型学習法によって生成された多

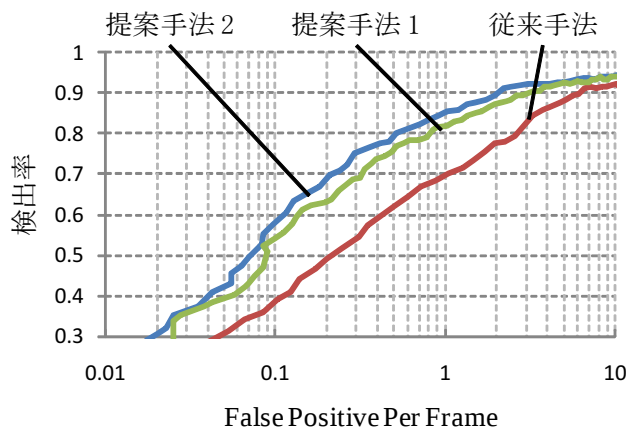


図 14 検出精度



図 15 提案手法 2 の検出結果

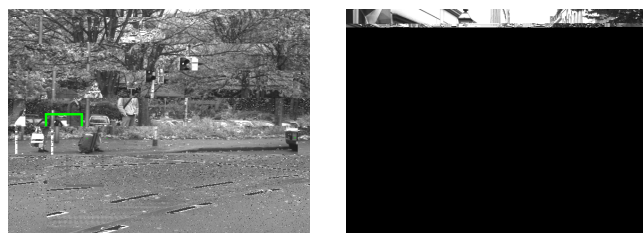


図 16 従来手法の検出結果

様な歩行者画像に対して姿勢毎に特化した識別器を構築し、それらを用いてマルチテンプレート型の識別器を構成することで、全体として姿勢変動に頑健な識別が実現できることがわかった。マルチテンプレート型の識別器を構成する各 2 クラス識別器は、ある歩行姿勢の歩行者のみ識別するため、これらを組み合わせることで歩行者の複雑な姿勢変化に対応できたといえる。

提案手法 2 と従来手法の実際の検出結果である図 15, 16 を見ると、従来手法では未検出や誤検出が発生しているのに対し、提案手法 2 では歩行者のみを正しく検出できていることがわかる。このことから検出精度が向上したと考えられる。

本実験では、学習用画像として 9,867 枚という決められた枚数の歩行者画像を生成し使用したが、生成枚数などの生成パラメータを変化させたときの検出精度の変化についても調べる必要があると考えられる。また、人手で収集した原画像のバリエーションを変化させた場合についても検出精度の変化を調べる必要がある。本手法では原画像をもとに歩行者画像を生成するため、原画像のバリエーションによって検出精度が変化すると考えられるためである。

4. む す び

本報告では、生成型学習法を用いた姿勢変化に頑健な歩行者検出手法を提案した。提案手法は、少数の歩行者画像から姿勢変化を考慮した多様な歩行者画像を多数生成し、姿勢を考慮したマルチテンプレート型の識別器を構築することで歩行姿勢の変化に頑健な歩行者検出を実現した。

今後の課題としては、照明条件を考慮した歩行者のテクスチャと背景の合成や、識別精度に大きく影響する形状クラスタリング方法の検討、生成パラメータの詳細な検討が挙げられる。

謝辞 日頃より熱心な御指導と御鞭撻を頂く名古屋大学村瀬研究室諸氏に深く感謝する。本研究の一部は、JST 戦略的創造研究推進事業 CREST および科学研究費補助金による。また、本研究では画像処理に MIST ライブラリ (<http://mist.murase.m.is.nagoya-u.ac.jp/>) を使用した。

文 献

- [1] N. Dalal, and B. Triggs, "Histograms of oriented gradients for human detection," Proceedings of 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol.1, pp.886–893, June 2005.
- [2] 村瀬洋, "画像認識のための生成型学習," 情報学論, vol.46, no.15 (CVIM 12), pp.35–42, Oct. 2005.
- [3] 道満恵介, 出口大輔, 高橋友和, 目加田慶人, 井手一郎, 村瀬洋, "色変動を考慮した生成型学習法による道路標識検出器の構築," 信学論 (D), vol.J93-D, no.8, pp.1375–1385, Aug. 2010.
- [4] M. Noda, T. Takahashi, D. Deguchi, I. Ide, H. Murase, Y. Kojima, and T. Naito, "Recognition of road markings from in-vehicle camera images by a generative learning method," Proceedings of the 11th IAPR Conference on Machine Vision Applications, no.15-5, pp.514–517, May 2009.
- [5] M. Enzweiler, and D.M. Gavrila, "A mixed generative-discriminative framework for pedestrian classification," Proceedings of 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.1–8, June 2008.
- [6] J. Marín, D. Vázquez, D. Gerónimo, and A.M. López, "Learning appearance in virtual scenarios for pedestrian detection," Proceedings of 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.137–144, June 2010.
- [7] 額綱直也, 山内悠嗣, 藤吉弘直, "人体シルエットの生成型追加学習による人検出の高精度化," 信学技報 PRMU2010-70, vol.110, no.188, pp.99–104, Sep. 2010.
- [8] D.M. Gavrila, and J. Giebel, "Virtual sample generation for template-based shape matching," Proceedings of 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol.1, pp.676–681, Dec. 2001.
- [9] T.F. Cootes, C.J. Taylor, D.H. Cooper, and J. Graham, "Active shape models. Their training and application," Computer Vision and Image Understanding, vol.61, pp.38–59, Jan. 1995.
- [10] DaimlerAG, "Daimler pedestrian detection benchmark data set," http://www.gavrila.net/Research/Pedestrian_Detection/Daimler_Pedestrian_Benchmarks/Daimler_Pedestrian_Detection_B/daimler_pedestrian_detection_b.html, Oct. 2008.
- [11] C.C. Chang, and C.J. Lin, "LibSVM A library for support vector machines," <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>.