

歩行者の見つけやすさ推定のための視覚探索特性を考慮した画像特徴の検討

○ 若山雅史 †, 道満恵介 †, 出口大輔 †, 井手一郎 †, 村瀬洋 †, 玉津幸政 ‡

○ Masashi WAKAYAMA †, Keisuke Doman †, Daisuke DEGUCHI †,
Ichiro IDE †, Hiroshi MURASE †, and Yukimasa TAMATSU ‡

†: 名古屋大学大学院情報科学研究科,

{mwakayama,kdoman}@murase.m.is.nagoya-u.ac.jp

{ddeguchi,ide,murase}@is.nagoya-u.ac.jp

‡: 株式会社デンソー

本報告では、ドライバにとって見つけにくい歩行者の情報のみを提示する運転支援システムの実現を目指し、歩行者の“見つけやすさ”を画像特徴から推定する手法を提案する。本報告では特に動画像を対象とし、オプティカルフローに基づく動き特徴、人の視覚探索の特性を考慮した特徴、その他見つけやすさに影響を及ぼすであろう特徴を複数検討した。そして、これらの特徴を歩行者を含む動画像から抽出し、SVR を利用した回帰により歩行者の見つけやすさを推定した。実験の結果、これらの特徴が歩行者の見つけやすさの推定に有効であることを確認した。

<キーワード> 歩行者, 見つけやすさ, 視覚探索, 顕著性, 車載カメラ映像

1. はじめに

近年、車載カメラを用いて歩行者を検出し、歩行者との衝突事故を防止する安全運転支援システムの実用化が期待されている。このようなシステムでは、衝突する危険性が高いと判断される歩行者を、音やディスプレイ等を用いてドライバに提示することが有効である^[1]。しかし、過度の情報提供はドライバの苛立ちや注意力低下を招くおそれがある^[2]。そのため、検出した歩行者情報を状況に応じて適切に取捨選択する技術が必要である。

そこで本研究では、運転時に見落としやすい歩行者のみをドライバに提示する運転支援システムの実現を目指し、車載カメラで撮影した動画像から歩行者の見つけやすさを推定する手法を提案する。

見つけやすい歩行者と見つけにくい歩行者の例を図1に示す。左の歩行者は中央付近に大きく映っており、また背景とのコントラストが高いため見つけやすいが、右の歩行者は複雑な背景に紛れて見つけることが難しい。また、左の歩行者は衝突の危険性があるが、見つけやすいため多くの場合警告する必要はない。逆に、右の歩行者が急に飛び出してきた場合、ドライバはこの歩行者に気づいていない可能



図1 見つけやすい／見つけにくい歩行者の例

性が高く、警告が必要である。このように、見つけやすさを推定できれば、見落としによる事故の危険を大きく低下させることが可能になると共に、ドライバの情報過多の防止にも貢献できる。

Engel らは、静止画を用いた歩行者の見つけやすさ推定手法を提案している^[3]。Engel らは、まず路上で撮影した静止画を被験者に瞬間提示し、提示後に被験者が歩行者位置を正しく入力できた割合を求め、歩行者の見つけやすさの目標値を計算した。そして、画像および画像に付与されたメタデータ（タグ付け・セグメンテーションされた領域群）等から抽出した特徴を用いて、回帰により見つけやすさを推定している。しかし、動画像を対象としておらず、

動きに関する特徴が検討されていない。また，“歩行者を見つける”というタスクは，視界から対象物体を探索する“視覚探索^[4]”に密接に関係し，視覚探索には視界中の妨害刺激が大きな影響を及ぼすことが分かっている。しかしながら，これについて明示的な検討はなされていない。

本報告では，Engel らの実験を動画画像へ拡張し，オプティカルフローを用いた動き特徴，および人間の視覚探索特性を考慮した特徴を新たに導入する。そして，SVR^[5]を用いた回帰により歩行者の見つけやすさを推定する手法を提案する。

以降 2 節で歩行者の見つけやすさ推定手法および推定に用いる特徴について述べる。そして，3 節で見つけやすさの目標値測定実験および見つけやすさ推定実験について述べる。4 節で実験結果に対して考察し，5 節にて本報告をむすぶ。

2. 歩行者の見つけやすさ推定

2.1 手法概要

本稿では，歩行者の見つけやすさ推定精度向上に寄与すると考えられる動き特徴，および人間の視覚探索の特性を考慮した特徴を検討する。上記に加え，様々な特徴を動画画像から抽出し，回帰を用いて見つけやすさの推定を行い，新たに追加した特徴が有効であることを確認する。回帰の学習段階では，各歩行者に対し見つけやすさに影響すると考えられる特徴を抽出し，それらと各歩行者の見つけやすさの目標値を用いて SVR を学習する。見つけやすさの目標値は被験者実験により獲得する。推定段階では，未知の歩行者に対し学習時と同様の特徴を抽出し，学習した SVR を用いて見つけやすさを推定する。

歩行者から特徴を抽出する流れを図 2 に示す。入力動画画像は，非常に短い区間を単位区間として切り出したものである。そのため，画像上での歩行者の位置はほとんど変化しない。入力動画画像をフレーム単位に分割し，各フレームから抽出される画像特徴と，複数フレームを用いて計算される動き特徴を歩行者毎に計算する。なお，歩行者位置は各フレームに対し既知であるとする。

歩行者の見つけやすさには様々な要素が影響すると考えられるが，本報告では特徴の持つ性質に基づいて大きく“動き特徴”，“歩行者 - 周辺コントラストに基づく特徴”，“視覚探索特性を考慮した特徴”，“歩行者のみに依存する特徴”，“その他の特徴”の 5 種類に分類した。実際に使用した特徴を表 1 に示す。説明の都合上，各特徴を式(1)のような記号で表

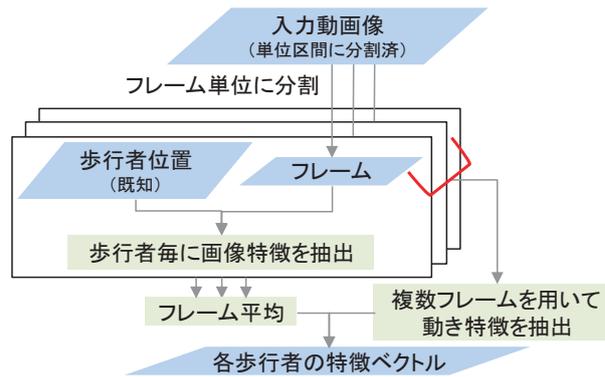


図 2 特徴抽出の流れ

表 1 見つけやすさ推定に使用した特徴

分類記号	特徴の略記	内容
<i>M</i>	flow	オプティカルフローを用いた動き特徴
<i>C</i>	bri	歩行者 - 周辺領域間の画像特徴のコントラスト（差）に基づく特徴
	color	
	edge	
<i>D</i>	freq	hist(x)は x のヒストグラム
	hist(color)	ヒストグラムの差としてヒストグラムインタセクションを使用
<i>D</i>	dist	視覚探索特性を考慮した特徴 画像中の妨害刺激 (distractor) を推定・抽出し，その数および歩行者との顕著度の差を使用
	num	
<i>E</i>	num	歩行者のみに依存する特徴 動画画像中の歩行者の数，面積，縦横比
	area	
	asp	
<i>O</i>	d(ped, c)	その他の特徴 d(x, y)は x と y のユークリッド距離 c は画像中心，n_ped は歩行者 ped に最も近い歩行者を意味する
	d(ped, n_ped)	
	bri(ped and c)	

現する。

$$F_a \quad (1)$$

ここで， F は分類記号であり， a は使用する特徴の略記号である。分類記号は以下のとおりである。

- (1) 動き特徴： M
- (2) 歩行者 - 周辺コントラストに基づく特徴： C
- (3) 視覚探索特性を考慮した特徴： D
- (4) 歩行者のみに依存する特徴： E
- (5) その他の特徴： O

特徴の略記号は，例えば色ならば color，色ヒストグラムならば hist(color)のように記述する。以降で各分類に属する特徴とその具体的な抽出手法について述べる。なお被験者実験は 3 節で述べる。

2.2 動き特徴

対象の動きにより見えに変化が生じるため，動いている物体の方が動いていない物体よりも見つけや

すいと考えられる。本報告では、画像の動き解析に広く用いられるオプティカルフローを用いて動き特徴を抽出する。

- M_{flow}

図 3(a)に示すように歩行者領域 p と歩行者周辺領域 b を定義する。動画画像中の各フレームから、隣り合う 2 フレームを用いて領域 p および領域 b におけるオプティカルフローを計算する。各領域でフローの方向ヒストグラムを作成し、そのヒストグラムインタセクションのフレーム平均を計算する。フローの方向は 360° を均等に 16 分割した。

2.3 歩行者 - 周辺コントラストに基づく特徴

道路照明における視認性評価では、視対象物と背景の輝度のコントラストを用いるのが一般的であり、コントラストが高いほど視認性が良い^[6]。また、視認性は見つけやすさに影響すると考えられる。そこで、本研究では輝度以外にも、色やエッジ強度などの画像特徴に関して歩行者とその周辺領域の差を計算し利用する。

- C_{bri} , C_{color} , C_{edge}

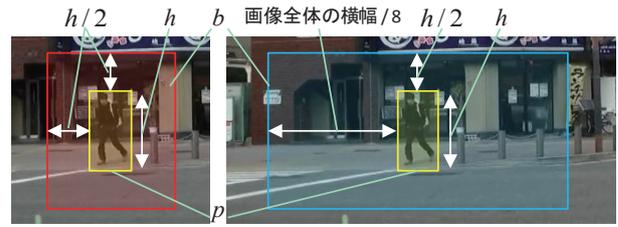
図 3(a)に示すように歩行者領域 p と歩行者周辺領域 b を定義する。領域 p と領域 b それぞれの平均輝度、平均色、平均エッジ強度の差の絶対値を計算する。なお、平均色の計算には L^*a^*b 表色系を使用し、差には L2 ノルムを用いた。エッジ強度の計算には Sobel フィルタを使用した。

- $C_{\text{hist}(\text{color})}$, $C_{\text{hist}(\text{edge_ori})}$

図 3(a)に示すように歩行者領域 p と歩行者周辺領域 b を定義する。領域 p と領域 b でそれぞれ色、および勾配方向ヒストグラムを求め、 p と b の間のヒストグラムインタセクションを計算する。色ヒストグラムは RGB の各成分を均等に 16 分割したビンを使用し、勾配方向ヒストグラムは 360° を均等に 16 分割したビンを使用した。

- C_{freq}

図 3(b)に示すように歩行者領域 p と歩行者周辺領域 b を定義する。また、説明のため領域 b に領域 p を加えた領域 $b'=b+p$ を導入する。画像全体をグレースケールに変換し、領域 b' にハミング窓をかけた画像と、 p より少し大きめの領域にハミング窓をかけた画像を用



(a) (b)以外での定義

(b) C_{freq} で用いた定義

図 3 歩行者領域 p と歩行者周辺領域 b の定義 (h は歩行者の高さ)

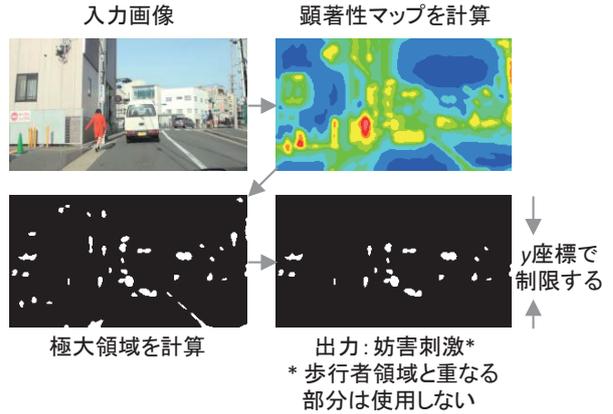


図 4 妨害刺激推定の流れ

意し、それぞれ領域 b' で切り取った画像を $I_H(b')$, $I_H(p)$ とする。 $I_H(b')$ および $I_H(p)$ に FFT を施し、両者の対応する周波数の振幅スペクトルの差の平均を計算する。

2.4 人間の視覚探索の特性を考慮した特徴

物体を探すタスクである視覚探索では、妨害刺激の存在が探索効率に影響する^[4]。より具体的には、妨害刺激の数が多いほど、また妨害刺激の持つ特徴と視対象物の持つ特徴が近いほど探索効率は下がる。この性質に着目し、我々はこれまでに顕著性マップ^[7]を用いて妨害刺激を推定する手法を提案し^[8]、歩行者の視認性評価に一定の効果があることを確認した。そこで、同様の手法で妨害刺激を推定し、その数等を利用する。

- D_{dist} , D_{num}

まず視覚探索の効率に影響する妨害刺激を顕著性マップを用いて推定する^[8]。その手順を図 4 に示す。推定した妨害刺激の数および歩行者領域 p の平均顕著度と妨害刺激の平均顕著度の差の総和を特徴として用いる。ここで、ドライバーから見た歩行者の位置は、水平方向の分散は大きい垂直方向の分散

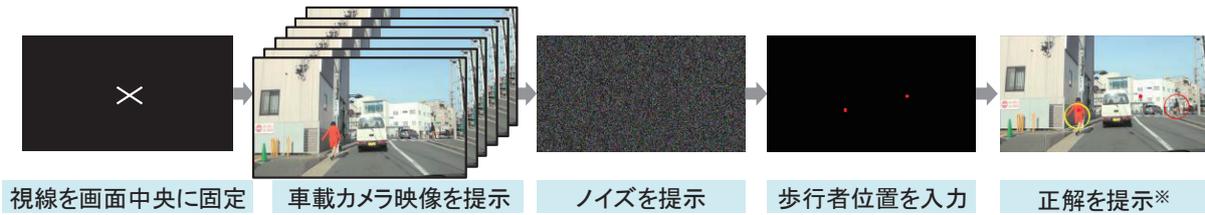


図5 見つけやすさの目標値測定のための被験者実験の流れ
(※正解の提示は実験全体の2割に対しランダムに実施)

は小さい^[10]. そのため, 歩行者を探す際に妨害刺激となりうる物体は垂直方向のある範囲に制約を受けると考えられる. そこで, 回帰の学習に用いた歩行者重心の y 座標が取りうる値を求め, その範囲内 (画像の高さを H として, およそ $0.26H < y < 0.6H$) に重心がある妨害刺激のみを抽出する.

2.5 歩行者のみに依存する特徴

- $E_{\text{num}}, E_{\text{area}}, E_{\text{asp}}$
画像中の歩行者数, 対象とする歩行者の面積および縦横比を計算する.

2.6 その他の特徴

上記以外の見つけやすさに影響すると考えられる特徴はこれに分類する.

- $O_{d(\text{ped}, c)}, O_{d(\text{ped}, n_{\text{ped}})}$
計算対象の歩行者と, 被験者実験で使用する視線固定位置 (本報告では常に画像中心) との距離, および最も近い歩行者との距離を L_2 ノルムで計算する. 動画中に歩行者が一人しか存在しない場合, $O_{d(\text{ped}, n_{\text{ped}})} = W$ (W は画像の横幅) とする.
- $O_{\text{bri}(\text{ped and } c)}$
被験者実験における視線固定位置 (画像中心) 周辺 15×15 ピクセルの領域と歩行者領域 p 間の平均輝度の差を計算する.

2.7 回帰に基づく見つけやすさ推定

見つけやすさの目標値を被験者実験を基に設定し, 表 1 の特徴を用いて SVM の回帰への応用である SVR^[5] のパラメータを学習する. 学習に際し, 全ての特徴は全歩行者に対し平均 0, 分散 1 となるように正規化する. また, カーネルには RBF を使用する. そして, 未知の歩行者から同様に特徴を抽出して正規化し, 学習した SVR を用いて見つけやすさを推定する. なお, SVR の学習及び推定には LIBSVM^[9] を用いる.

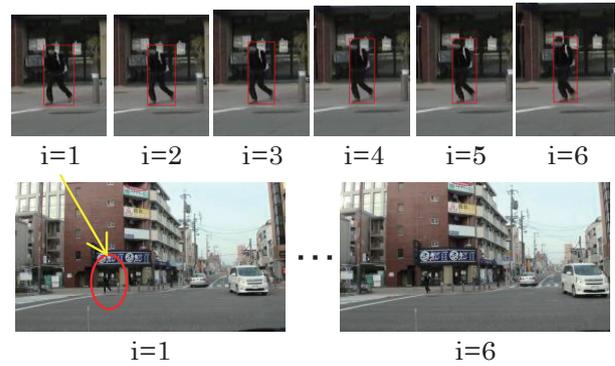


図6 被験者に提示した車載カメラ映像の例
(i はフレーム番号)

3. 実験

3.1 見つけやすさの目標値測定実験

Engel らの静止画に対する実験を動画像へ拡張し, 歩行者の見つけやすさの目標値を測定した. 図 5 に被験者実験の流れを示す. 本実験では, ディスプレイに瞬間提示される動画像を被験者に見てもらい, 画面が暗転した後に歩行者が存在した位置を入力させた. データセットは昼間に車載カメラで撮影した 1280×720 ピクセルの動画像から, 6 フレームを切り出して作成した全 585 系列を使用した (例: 図 6). フレームレートは 30fps であり, 動画の提示時間は約 200ms である. また, 一つの系列に含まれる歩行者数数は 0~4 人とした. 歩行者数別の系列数は 0 人: 93 系列, 1 人: 242 系列, 2 人: 154 系列, 3 人: 70 系列, 4 人: 26 系列であり, 動画に含まれる歩行者は全部で 864 人であった.

20 歳台から 50 歳台の男女 7 人に対し本実験を行い, 実験結果から次式の Engel らの定義^[3]に従って歩行者の見つけやすさ D (歩行者) の目標値を測定した.

$$D(\text{歩行者}) = P(\text{入力位置が正解領域内に存在}) \quad (2)$$

ここで, $P(A)$ は A が生じる確率を表す. 歩行者位置の正解領域は各フレームにおける歩行者の重心から半径 100 ピクセル以内 (いずれかのフレーム中の

正解領域に収まっていれば正解)とした。また、順序効果を打ち消すため、動画像の提示順は被験者毎にランダムに並べ替えた。さらに、疲労の効果を打ち消すため、図5の試行を100回繰り返すごとに被験者を休憩させた。

実験の結果、見つけやすさ D 毎の歩行者数は図7のようになった。 $D=1$ となった歩行者が最も多く、全体の約35%を占めているが、それ以外はおよそ均等に分布した。Engelらの静止画に対する実験においてもほぼ同様の結果を得ているが、動画像に対する実験では $D=1$ となった歩行者の割合がより高くなった。この原因として、提示時間が Engel らの100ms に対して200ms と長くなったこと、動画像が持つ動きの情報が歩行者をより見つけやすくしたこと、などが考えられる。

3.2 画像特徴の有効性に関する評価実験

本報告の目的の一つは、どの特徴が歩行者の見つけやすさに寄与するかを明らかにすることである。そこで、以下の処理を行った。

- (1) 特徴の組合わせ集合 X に表1の特徴を代入
- (2) X から i 番目の特徴を除外した特徴集合 $X_{-i} = X - \{x_i \in X\}$ を用いて SVR を学習し、その性能 R_{-i}^2 を最大にするインデックス $m_{-i} = \arg \max_i R_{-i}^2$ を求める
- (3) $X = X_{-m_{-i}}$ とする
- (4) $|X| > 2$ なら(2)へ、そうでなければ終了。

ここで、 R^2 は決定係数 (または寄与率) であり、回帰の当てはまり具合を表現する評価値である。これは式(3)で表される。

$$R^2 \equiv 1 - \frac{\sum_i (y_i - f_i)^2}{\sum_i (y_i - \bar{y})^2} \quad (3)$$

y_i は標本値、 \bar{y} は標本平均、 f_i は推定値である。ここで、(1)~(4)の処理の流れの中で、処理の後半まで除外されなかった特徴ほど見つけやすさに寄与すると考えられる。SVR の学習および評価では、全864人の歩行者をランダムに5つのグループに分け、5-fold cross validation により平均決定係数と平均絶対誤差を計算した。

実験の結果を表2に示す。本報告で新たに提案した3つの特徴、(M_{flow} , D_{num} , D_{dist}) は、全15個の特徴からそれぞれ8/12/5番目に除外された。また、平均決定係数の値は特徴数が11次元のときに最高の0.351を示した。これは上記3つの特徴が除

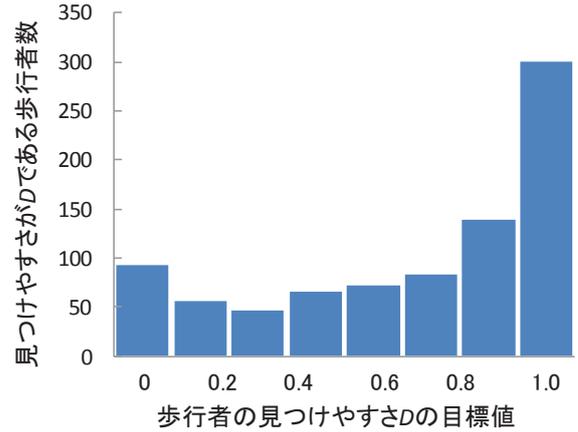


図7 見つけやすさの目標値毎の歩行者数ヒストグラム

表2 除外された特徴と見つけやすさ推定精度

特徴数 (除外後)	除外された 特徴	推定精度	
		平均 R^2	平均絶対誤差
15	-	0.304	0.236
14	$C_{hist(edge_ori)}$	0.321	0.233
13	E_{asp}	0.341	0.229
12	C_{bri}	0.346	0.229
11	C_{edge}	0.351	0.229
10	D_{dist}	0.351	0.229
9	E_{num}	0.342	0.227
8	C_{freq}	0.329	0.230
7	M_{flow}	0.325	0.228
6	$O_{bri(ped\ and\ c)}$	0.275	0.234
5	$C_{hist(color)}$	0.276	0.233
4	C_{color}	0.248	0.237
3	D_{num}	0.238	0.238
2	$O_{d(ped,\ n_ped)}$	0.148	0.246

最後まで残った特徴 $O_{d(ped, c)}$, E_{area}

外されるより前の段階である。平均絶対誤差は約0.23付近を推移した。

4. 考察

画像特徴の有効性に関して考察する。評価実験の結果、本報告で提案した3つの特徴はSVRの決定係数が最も高くなった際の特徴の組み合わせに含まれており、見つけやすさ推定に一定の効果があると考えられる。表2の結果から、特に D_{num} 、すなわち妨害刺激の数に基づく特徴が処理の後半まで除外されず、推定に有効であったといえる。このことから、人間の視覚探索特性を考慮することは見つけやすさ推定に重要であると考えられる。しかしながら、 D_{dist} 、すなわち妨害刺激の顕著度に基づく特徴はさほど寄与しなかったため、特徴の設計方法の更なる検討が必要である。

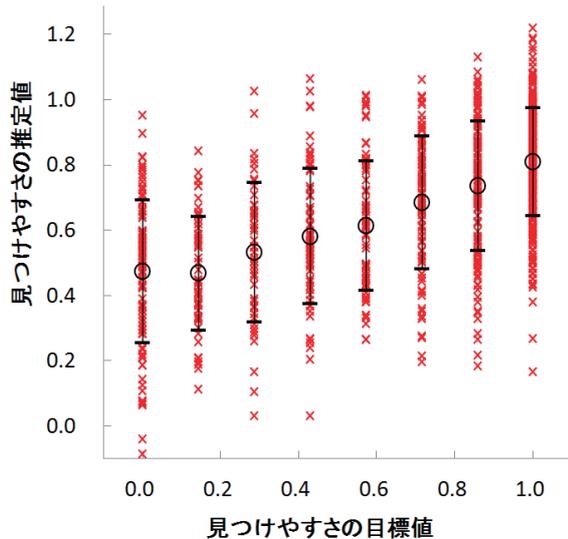


図 8 学習した SVR による見つけやすさの予測結果

表 2 において、平均決定係数の最高値を示した際の特徴 (C_{edge} が除外された後の 11 次元) を用いて見つけやすさを推定した結果を図 8 に示す。横軸が見つけやすさの目標値、縦軸が推定値である。また、各プロットは推定された歩行者のサンプル点であり、円のプロットおよび直線は、目標値が同一であるプロットの推定値の平均と標準偏差を表す。推定値の平均が 0.6 付近に寄っており、推定精度は高いとは言えない。そのため、今後見つけやすさをより良く表現する特徴の検討が必要である。

また、回帰は推定値と目標値の誤差を最小化するという性質を持つため、学習データ内で見つけやすさの目標値が偏っている場合推定に悪影響を及ぼす。例えば、見つけやすさが 1.0 である歩行者が多すぎる場合、推定値も 1.0 の方にバイアスがかかる。そのため、学習に用いる歩行者の選択や、見つけやすさの目標値測定実験の方法の検討が必要である。

5. むすび

車載カメラで撮影した動画から、歩行者の見つけやすさを推定する手法を提案した。動き特徴と視覚探索特性を考慮した特徴 (妨害刺激の数) が見つけやすさに一定量寄与することを確認した。今後の課題として、シーンコンテキストの影響の調査や、より長い動画における歩行者の見つけやすさ推定が挙げられる。

謝辞

日頃より研究のご指導を頂いている名古屋大学村

瀬研究室諸氏に深く感謝する。本研究の一部は、JS T 戦略的創造研究推進事業 CREST, 科学研究費補助金, および柏森情報科学振興財団研究助成による。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。

参考文献

- [1] トヨタ自動車: “プリクラッシュセーフティシステム,” (http://www.toyota.co.jp/jpn/tech/safety/technology/pre_crash/)
- [2] 北村康弘, 畑山俊輝, “自動車運転時の聴覚情報への対応課題がドライバーに与える影響-生理的・行動的变化から-,” 国際交通安全学会誌, vol.30, no.3, pp.327-332, Sept. 2005.
- [3] D. Engel et al., “Pedestrian Detectability: Predicting Human Perception Performance with Machine Vision,” Proc. 2011 IEEE Intelligent Vehicles Symposium, pp.429-435, June 2011.
- [4] J.M. Wolfe, “Visual search,” in Attention, ed. H. Pashler, vol.1, pp.13-73, University College London Press, 1998.
- [5] D. Basak et al., “Support Vector Regression,” Neural Information Processing - Letters and Reviews, vol.11, no.10, pp.203-224, Oct. 2007.
- [6] 魚住拓司, 上野朋子, 川上幸二, “道路照明のビジビリティレベルに関する研究-視対象物の視認性を向上させる器具配光のシミュレーション-,” 照明学会全国大会講演論文集, p.161, Sept. 2001.
- [7] L. Itti, C. Koch, and E. Niebur, “A model of saliencybased visual attention for rapid scene analysis,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol.20, no.11, pp.1254-1259, Nov. 1998.
- [8] 若山雅史, 道満恵介, 出口大輔, 井手一郎, 村瀬洋, 玉津幸政, “大局的特徴として顕著性を利用した歩行者の視認性推定,” 画像の認識・理解シンポジウム (MIRU2011)講演予稿集, pp.1597-1603, July 2011.
- [9] C.-C. Chang and C.-J. Lin, “LIBSVM -- a library for support vector machines,” (<http://www.csie.ntu.edu.tw/~cjlin/libsvm/>)
- [10] A. Torralba, A. Oliva, M. Castelhano, and J. Henderson, “Contextual guidance of eye movements and attention in real-world scenes: The role of global features on object search,” Psychological Review, vol.113, no.4, pp.766-786, Oct. 2006.