Vol.2022-CVIM-230 No.30
2022/5/12

道路死角データセット

石崎 慎弥[1]　長谷川 浩太郎[1]　福田 太一[1]　延原 章平[1]　西野 恒[1]

概要：安全な自動運転の実現には見えていない箇所をコンピュータが認識することが必要である．しかし，死角とは本来人やモノに遮蔽された三次元空間であるため，一枚の画像から直接求めることはできない．そこで本研究では死角を，現在見えないが将来見える道路領域と定義する．これにより，カメラパラメータが既知の任意の動画から各フレーム内の死角をピクセル単位で示したデータセットを作成できる．本研究では実際に自動運転用データセットの動画からフレームごとの死角を示した道路死角データセットを作成した．作成したデータセットを用いることで画像一枚から死角を推定するネットワークを訓練することができる．

1. 安全な自動運転実現のための死角検出

安全な自動運転の実現には，見えているものから見えていない箇所の存在を認識し飛び出しの危険を検出することが不可欠である．多数のセンサを搭載した車に比べ，センサ数では及ばない人間が，より明確に周囲の危険を理解できている理由として，人間は見えていない箇所があることを知っているという点が挙げられる．コンピュータにも人間と同じように明示的に見えていない箇所の存在を認識させることが本研究の動機である．

死角とは本来運転者の視点から見えない三次元空間すべてを指すが，このような空間を全て明示するためには道路シーン全体の詳細な三次元復元が必要であり，これを単一の画像から行うことはできない．そこで，本研究では，死角を現在見えていないが車が進むにつれて見えるようになる道路領域と定義し，その自動的な計算手法を提案する．これによりカメラパラメータが既知の任意の車載動画について死角を求めることが出来る．提案手法の入力は運転シーンの車載動画，出力は各キーフレームの死角を示すバイナリマスクである．まず車載動画に，自己位置推定（SLAM），単眼深度推定，セマンティックセグメンテーションの手法を適用し，移動軌跡，各フレームの深度及び道路領域を得る．移動軌跡から二つの時刻の間でカメラがどう動き，どう回転したかを示す相対的なカメラ姿勢が求められる．これらの情報から，先の時刻での道路領域を現在の時刻に投影することで，現在の画像内でその領域が見える場所を計算することができる．得られた道路領域の重ね合わせと現在の時刻から見えている道路領域の情報を比較することで死角を得ることができる．

本研究では車載映像のデータセットとして KITTI データセット [1]，BDD データセット [2]，TITAN データセット [3] を用いた．死角は 5 秒以内に見える道路領域であるとし，それぞれ 51，62，118 の動画から 5fps で 51 分，34 分，12 分の動画の死角を示すバイナリマスクを得た．得られた死角を検証するため，以下の評価実験を行った．KITTI データセットの LiDAR センサによる三次元点群情報をもとに疎な三次元世界を復元する．ある一定の高さより低い箇所を道路領域と仮定し，本手法同様に 5 秒間の道路領域を集約し，推定された死角と Precision で評価した．

本研究で得られたデータセットは画像一枚に対し，画像内の死角を示すバイナリマスクを提供する．これは二次元から二次元への変換であり，本データセットを用いることで画像一枚からそこに映る死角を推定するネットワークを訓練することが出来る．

2. 関連研究

これまでの手法では，任意の車載動画で死角を求めることは難しかった．関連する手法で複数のカメラや視点変化により周囲の状況を理解しようとする研究は存在する．

見えない領域を明示する方法として，Barnum らは，交差点などに設置された監視カメラなどから映像を転送することで，死角にある見えない物体を透視的に見せる方法を提案している [4]．彼らは，視点となる可動カメラと遮蔽物を映す固定カメラの位置関係を背景などの情報から推測する．この位置関係からホモグラフィー変換を通じて固定カメラの映像を可動カメラに投影する．彼らのシステムは 5Hz で駆動し，飛び出そうとするものや人があればすぐに

[1] 京都大学大学院情報学研究科

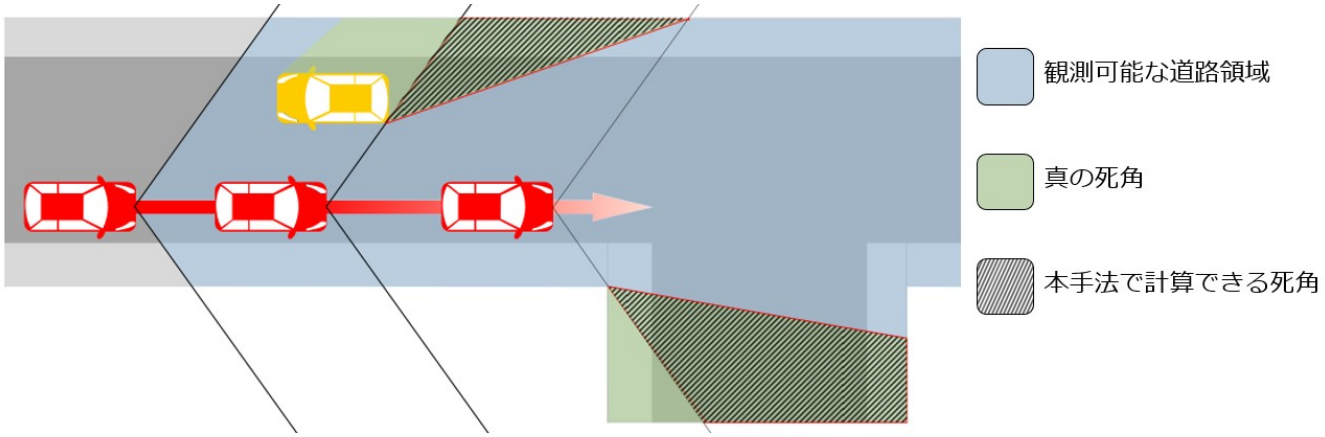

図 1 本研究では死角を，現在見えないが将来見える道路領域と定義する．真の死角の中にいる歩行者や自転車が移動中の車と接触するためには，車から見える道路領域もしくは，本研究で定義する死角領域を必ず通らなければならないことが分かる．

知ることができ，事故を未然に防ぐことができる一方，監視カメラが設置されていない場所では動作しない．

また，鳥瞰図のような俯瞰した視点で道路環境を復元する研究も存在する．Zhu らは敵対的生成ネットワークを用いて正面図と鳥瞰図の 90 度の視点変換において歪みを少なくすることに成功した [5]．Mani らは文献 [6] で Yang らは文献 [7] で他にも多層ニューラルネットを用いることで，道路領域内の車の占有領域を推定する手法 [7] や見えない領域を推測する手法 [6] で運転手の視点から鳥瞰図に変換する研究がなされている．

Watson らは一枚の RGB 画像から，道路平面の中で通行可能な領域を推測するモデルを提案した [8]．彼らのモデルでは，画像から見えている通行可能領域だけでなく，画像内のオブジェクトの後方の通行可能領域も推測することができる．また，彼らは提案モデルの教師となる学習データセット作成のために，ステレオ画像から通行可能な領域を計算する手法を提案している．

3. 提案手法

3.1 死角の定義

安全な自動運転の実現のためには，見えていない箇所，つまり死角がどこにあるかを認識することが欠かせない．しかし，死角がどこにあるかということは表面的な情報ではないので，運転シーンにおける単一の画像から，死角がどこにあるかを直接計算することは難しい．そこで本研究では図 1 のように，死角を現在は見えないが将来見える道路領域と定義する．

同図から明らかなように，この定義では検出できない死角も確かに存在する．しかし，そのような領域内にいる歩行者や自転車が進行中の車と接触するためには，車から見える道路，もしくは本研究で定義する死角を通る必要がある．したがって本研究で定義した死角を求めることは，飛び出しの可能性がある領域を示すことに対応する．

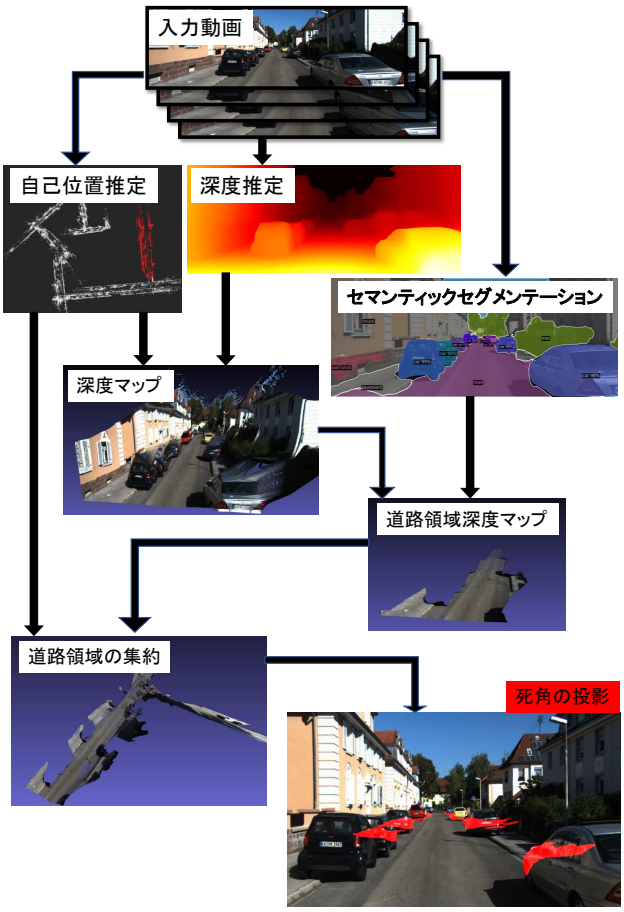

図 2 入力は車載動画，出力は各キーフレームの死角を表すバイナリマスクである．先の時間に映る道路領域を，深度情報を用いて現在のフレームに投影し，現在のフレームの道路領域を差し引くことで死角を得る．道路領域，深度情報，投影に必要なカメラの外部パラメータはそれぞれセマンティックセグメンテーション，深度推定，自己位置推定を適用して得られる．

3.2 手法概要

図 2 に提案手法の概略を示す．入力は車載動画，出力は各フレームから T 秒以内に見える死角を示すバイナリマス

クである．

3.1 節で定義したしたように，死角は各フレームの道路領域の一部であるから，まずフレーム I_t から T 秒以内に見える道路領域をセマンティックセグメンテーションを適用して求める．フレーム I_{t+1} から I_T においてそのフレームで見える道路領域は，Watson らのデータセット作成の手法 [8] を用いて，フレーム I_t に投影することができる．道路領域を投影するにあたって，カメラの内部パラメータは既知であるとして，カメラのエゴモーション及び各フレームの道路領域の密な深度マップが必要になる．それぞれ Visual SLAM，単眼深度推定を用いてこれらを求める．Visual SLAM で得られる特徴点の三次元点群は疎な環境マップである一方，単眼深度推定で出力される深度は相対的な深度を表した深度の逆数である．これらを線形回帰で補正することにより密な深度マップを得る．投影された道路領域全体と，フレーム I_t から見える道路領域の論理否定の和を取って，死角を得る．

3.3 密な深度マップの作成

本研究における死角は，フレーム I_t において時刻 t では見えないが，T 秒以内に見える道路領域に対応するピクセル $x \in \Omega_t^T$ である．本手法の目標は，ピクセル x の集合としての死角を画像内のバイナリマスク $\omega(x,t;T) : \mathbb{R}^2 \times \mathbb{R} \mapsto \{0,1\}$ として求めることである．

Visual SLAM は，周囲の物体の特徴点を画像内から抽出し，三次元点群として登録することで疎な環境マップを得る．しかし，死角はピクセル単位で死角かどうか判断される必要があり，また道路領域を別のフレームに投影する際にもピクセルごとの深度情報が必要になる．一方，単眼深度推定で出力される深度は，スケール，バイアスが不定な相対深度であり，深度の逆数である．密な絶対深度マップを得るため，単眼深度推定の結果と Visual SLAM の疎な三次元点群を組み合わせる．

自己位置推定で得られた特徴点の三次元座標をそれぞれのフレームに投影する座標変換は透視投影変換と呼ばれ、画像座標、ワールド座標、カメラの内部パラメータ、外部パラメータを $(u,v,1)^{\mathrm{T}}$、$(X_w,Y_w,Z_w)^{\mathrm{T}}$、$\begin{pmatrix} \frac{f}{\delta_u} & & c_u \\ & \frac{f}{\delta_v} & c_v \\ & & 1 \end{pmatrix}$、$\begin{pmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_1 \end{pmatrix}$ として

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} \sim \begin{pmatrix} \frac{f}{\delta_u} & & c_u \\ & \frac{f}{\delta_v} & c_v \\ & & 1 \end{pmatrix} \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_1 \end{pmatrix} \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix} \tag{1}$$

図 3 自己位置推定で推定で得られる疎な環境マップと深度推定で得られる密な相対深度マップを組み合わせると画像全体の密な深度マップが得られる．カメラからは見えない車の裏などは情報を復元できない．

のように表せる。つまり、カメラの内部パラメータを既知として、ワールド座標及び外部パラメータを Visual SLAM で求めて、透視投影変換を行い元の画像上に投影する。

次に単眼深度推定で出力されたバイアスとスケールが不定な深度マップを補正する方法について述べる．投影された特徴点に対応するカメラ座標の Z 座標はその点のカメラからの奥行き，つまり深度に対応する．i 番目の特徴点 f_i に対し，特徴点深度を Z_{f_i}，その特徴点の画像座標 (u_i,v_i) における単眼深度推定の出力結果を $d_{inv}(u_i,v_i)$ とする．これらの集合を $\mathbb{Z}_{f_i}$，$\mathbb{D}_{inv}$ として，最小二乗法により $\mathbb{Z}_{f_i}$ を $\mathbb{D}_{inv}$ で線形回帰する．これにより，フレーム内で任意の画像座標での深度を環境マップと同じスケールの深度に変換することができる．図 3 にある入力画像を密な深度マップに変換した結果を示す．

3.4 死角領域の推定

本節では，得られた各フレームの深度マップから 3.1 節で定義した死角領域を推定する方法について述べる．

各フレームにおいて，3.3 節で得た深度マップにセマンテックセグメンテーションの道路領域の推定結果でマスク

処理をすると，道路領域の深度マップを得る．

フレーム I_t における道路領域を $r(x,t)$ とすると，$r(x,t+1) \sim r(x,t+T)$ を I_t に投影する必要がある．道路領域を別のフレームに投影する方法は Watson ら [8] のデータセット作成時の方法を用いた．彼らは 2 つのフレームの外部パラメータから相対的な外部パラメータを計算し，道路領域の深度マップを相対的な外部パラメータを用いて投影している．この手法で I_{t+1} 上の $r(x,t+1)$ は I_t 上の $r'(x,t+1;t)$ として投影される．投影された $r'(x,t+1;t) \sim r'(x,t+T;t)$ を集約して T 秒以内に見える道路領域全体を得る．つまり，I_t において，T 秒以内に見える道路領域全体 $s(x,t;T)$ は

$$s(x,t;T) = \quad r(x,t) \vee r'(x,t+1;t) \vee \cdots \vee r'(x,t+T;t) \tag{2}$$

と表せる．$A \vee B$ は A と B の論理和を表す．ここから，現在見えている道路領域 $r(x,t)$ を取り除くことで，現在見えないが T 秒以内に見える死角を得る．すなわち，死角領域 $\omega(x,t;T)$ は

$$\omega(x,t;T) = s(x,t;T) \wedge \bar{r}(x,t) \tag{3}$$

と表せる．$A \wedge B$ は A と B の論理積を表し，$\bar{A}$ は A の論理否定を表す．これにより，3.1 節で定義した死角を求めることができた．

3.5 深度比較による誤差の除去

本手法で計算された死角を観察すると道路の縁が死角として推定される場合がある．これはセマンティックセグメンテーションの誤差によるものであり，あるフレームでは道路領域ではないと判断されているが，先の時刻のフレームでは道路であると出力されているような場所がこれにあたる．このような箇所は，本研究の死角の定義に該当するが，本来意図している死角ではない．

これらの領域を推定された死角からフレーム I_t における道路領域のカメラ座標を用いて取り除く．つまり，死角として投影されたピクセルのカメラ座標系の Z 座標（死角深度）と，同じピクセルのフレーム I_t から見える深度（画像深度）を比較する．もし，推定された死角領域が本当に死角であった場合，画像深度と死角深度は異なるが，推定結果が誤っていた場合はこれらはは等しくなる．このことを利用して，死角深度が画像深度と同程度あれば死角から除去することでより正確な死角を得られる．死角深度を $d_a(x)$，画像深度を $d(x)$ とすると，

$$\omega(x,t;T) = \begin{cases} 1 & (|d(x) - d_a(x)| \geq l_d) \\ 0 & (|d(x) - d_a(x)| < l_d) \end{cases} \tag{4}$$

のように，閾値 l_d を決めて修正を行った．

また，影やタイヤなどにより作られる小さな死角領域も存在する．本研究の目的は飛び出しの可能性がある死角領域を検出することであり，このような死角は特段の注意が必要な死角ではない．そのため，領域全体が合計 100 ピクセル以下になるような推定結果は死角から除去した．

3.6 提案手法の適用範囲

本手法を用いて RGB 動画から死角のバイナリマスクを生成できるのは，自己位置推定や深度推定，セマンティックセグメンテーションなどの既存の手法が正しく動作する場合である．つまり本手法の限界は既存手法の限界と一致する．

自己位置推定はカメラパラメータを既知として動作する．つまり，動画が撮影されたカメラの内部パラメータを事前に知っておく必要がある．また，Visual SLAM は各フレームに写り込んだ物体の中から特徴点を取り環境マップを作成するが，特徴点が取られた物体は静止していることが望ましい．特に車載映像において，例えば動画を撮影している車に対し斜め前方で並走している車がある場合，車同士の相対的な位置関係は変わらない．つまり，画像における並走車の位置がほとんど変わらないことを意味するが，そのような特徴点は無限遠に存在すると推定されてしまう．これは密な深度マップを作成する上で精度を下げる原因となりうる．特に，高速道路や大きな幹線道路など斜め前方の左右ともに並走車両がある場合は正しく動作しない．しかし，並走を続ける車の影から突然歩行者や車が飛び出してくる，ということは考えられにくいので，このような状況の死角を作成する必要性は低い．本研究の目的は，歩行者や車が飛び出してくるような死角をラベル付けしたデータセットを作成することである．死角が存在するような複雑な道路状況になればなるほど，自己位置推定に有利な特徴点を多く取ることができる．

単眼深度推定やセマンティックセグメンテーションは RGB 画像のみを入力とするので，特定の事前知識を必要としない．これらの手法が正しく動作しない場合は以下のような場合が考えられる．

(1) 車やカメラの揺れなどにより撮影動画が大きくぶれている場合
(2) 動画が光などにより白飛びするもしくは，夜の暗い道路など周囲の状況が正しく認識できない場合
(3) 路面や周囲の環境がが雪に覆われている場合

1 については，この場合は自己位置推定もうまく動作しないことが考えられるが，ブレが長い時間に渡っているならそれはそもそも動画の撮影に失敗したと言える．その時間についての死角は他のどのような工夫を用いても計算できない．2 について，これも自己位置推定が正しく動作しないと考えられる．特に夜の運転は人間であっても注意深く運転すべきであり，自動運転技術にとって今後の課題とな

る．3 について，これは深度推定及び道路領域の推定が正しくできないと考えられる．

以上から，本手法が適用できるのは死角が存在するような日常的な道路を撮影した動画全般であると言える．一部な特殊な状況を除き，カメラパラメータの分かるあらゆる車載動画から死角を求めることができるのは本提案手法の大きな強みである．

4. 評価実験

4.1 データセット

死角ラベルの作成にあたっては, KITTI データセット [1], BDD データセット [2]，TITAN データセット [3] の動画を用いた．KITTI データセットはドイツのカールスルーエ工科大学とアメリカ・シカゴにある豊田工業大学シカゴ校のチームが作成した，自動運転のための画像データセットである．セマンティックセグメンテーションやオプティカルフローなどの研究分野においても訓練データやテストデータとして用いられる，自動運転研究の最も一般的なデータセットの一つと言える．BDD データセットは，ニューヨークやバークレーなどの都市での走行映像を 10 万本集めたデータセットである．他の車載動画のデータにセットに比べて，時間や天候などの多様性に富む．TITAN データセットは，東京都内において撮影された道路シーンが複雑に変化するデータセットである．他の車載動画のデータにセットに比べて，狭い道路でのシーンや歩行者と並走するシーンが多く，潜在的な危険が最も多いシーンを含むデータセットの一つである．

各データセットの動画に自己位置推定の手法を適用し，特徴点の三次元座標とキーフレームのエゴモーションを得る．自己位置推定では，数フレームに一枚の割合でキーフレームを抜きだし，エゴモーションを推定する．これは空間的なサンプリングであり，サンプリング周波数は動画やシーンによって異なる．

本研究ではあるフレームから 5 秒以内に見える死角を計算した．運転中の車が止まるための停止距離には空走距離と制動距離があり，一般道を走る車は概ね 5 秒以内に止まれる状態にあるため，5 秒以内に見える道路領域は飛び出した人や物が車と接触する可能性がある領域である．

自己位置推定には openvslam[9]，単眼深度推定には MiDaS[10]，セマンティックセグメンテーションには detectron2[11] を用いた．セマンティックセグメンテーションの道路領域は detectron2[11] の出力結果のうち，road と pavement と出力されたものの論理和と取った．車載動画のデータセットは，KITTI データセット [1]，BDD データセット [2]，TITAN データセット [3] のうち，それぞれ 51, 62,118 の動画から，5fps での 51 分，34 分，12 分の動画の道路死角マスクを得た．図 4 に作成したデータセットの一部を示す．

4.2 エゴモーションの定性的評価

本手法において自己位置推定の精度は密な深度マップの作成，道路領域の投影の精度に影響するため，本手法の正確性を示す重要な指標の一つである．

得られた外部パラメータと KITTI データセットに含まれる LiDAR センサの三次元点群を用いてワールド座標系での三次元世界を復元する．LiDAR センサで得られる三次元点群は LiDAR センサを中心としたカメラ座標系で表現されている。あるフレームにおいて三次元点群のカメラ座標系，ワールド座標系，外部パラメータの回転行列，並進ベクトルをそれぞれ $\tilde{\boldsymbol{X}}_c$，$\tilde{\boldsymbol{X}}_w$，$\boldsymbol{R}$，$\boldsymbol{t}$ とすると，カメラ座標系からワールド座標系への投影は

$$\tilde{\boldsymbol{X}}_w = \boldsymbol{R}^{-1}(\tilde{\boldsymbol{X}}_c - \boldsymbol{t}) \tag{5}$$

と表される．Visual SLAM で取得した並進ベクトルのスケールは LiDAR センサで取得した三次元点群のスケールとは一致しない．そこで，3.3 で述べたような方法を用いてこのスケールを求める．つまり，SLAM から得られた特徴点と LiDAR センサで得られた特徴点をそれぞれ画像座標上に投影し，近い点同士の深度で線形回帰を行う．ここで得られた傾きを並進ベクトルにかけて，Visual SLAM で得られた並進ベクトルのスケールを LiDAR センサで得られた三次元点群のスケールに合わせる．本研究では，死角を 5 秒以内に見える道路領域のうち現在見えないもの，と定義した．そのため，あるフレームに対し 5 秒後のフレームに対応するそれぞれの LiDAR 点群を投影する．

もし，得られたカメラの外部パラメータが正しければ家の壁や停車中の車など，その動画全体で静止している物体は重なって見える．逆に，外部パラメータの誤差が大きい場合，静止しているはずの物体上の点群は時間の経過とともに，異なる場所に投影される．図 5 に結果を示す．元のフレームの LiDAR 点を水色で，5 秒後先の LiDAR 点を黄緑でプロットしている． 図より，静止している物体は同じ場所に重なって投影されている．このことから，死角を計算する時間内において，外部パラメータは適切に推定できていると考えられる．

4.3 死角領域の検証

本手法により得られた死角の検証を行うため，以下のような実験を行った．死角を直接求めることは難しく，また死角の真値と呼べるものも存在しない．そのため，今回は KITTI データセット中の LiDAR 点群を利用する．この LiDAR センサで得られた点群とカメラポーズを用いることで道路領域を投影する．投影された道路領域のうち，現在のフレームから見えている道路領域を差し引くことで死角を得た．LiDAR センサで取得した点群のうち，センサよりも 1.5 メートル以下の点群を道路領域であるとした．

得られた三次元点群は道路領域上の点であるが，求めた

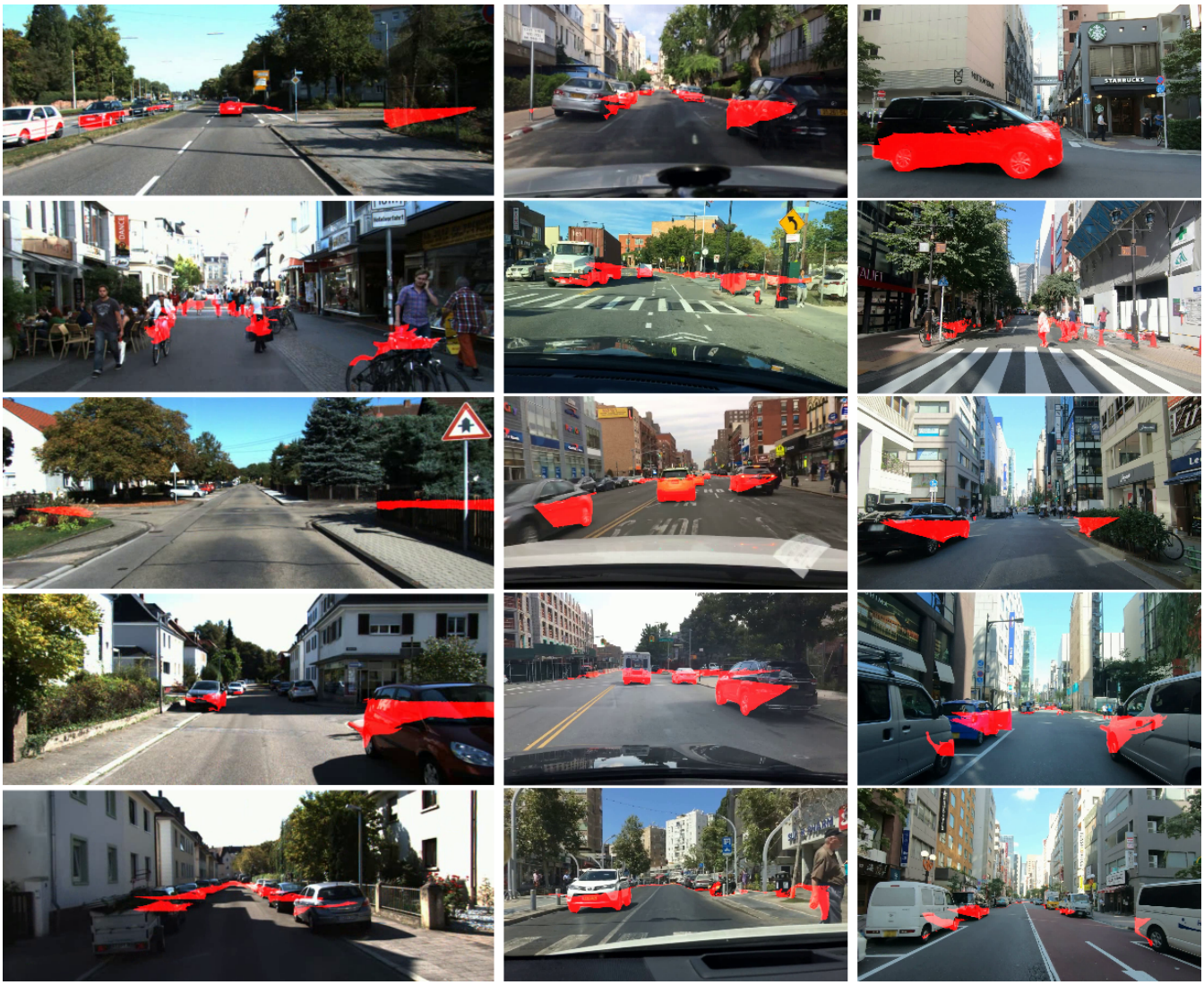

図 4 作成したデータセットの例．左から列ごとに KITTI データセット，BDD データセット，TITAN データセット．元画像に対し，死角のバイナリマスクを赤く重ねている．

い道路領域の全てのピクセルを得られたわけではない．この原因は二つある．第一に，どの時間からも LiDAR センサから見えない点が存在することがある．例えば，進行中の道の左側に車が停車していたとする．この場合は，停車中の車の左側には見えない部分が存在する．第二に，LiDAR センサの分解能の問題により，領域としては LiDAR センサから見えているのに，全てのピクセルを取得できない場合がある．第一の問題に対しては対処できない．第二の問題に対しては，二値画像に対する処理の一つであるモルフォロジー変換のクロージングを施すことにより対処した．

4.4 検証結果

本手法で求めた死角を推定結果，LiDAR で求めた死角を真値をして計算した．KITTI データセットのうち一つの動画を選び各キーフレームに対し，本手法を用いた死角と LiDAR 点を用いた死角を計算した．

本手法の死角の定義を考えてみれば，前方に向かって進む動画において LiDAR 点よりも得られる死角が少ないことは明らかである．通り過ぎた停車中の車近辺の道路領域を全て得ることは難しく，十字路を左に曲がる場合は曲がらなかった右側の死角を得ることはできない．このことから，評価指標は IoU や Recall よりも Precision の方が重要だと言える．

図 6，図 7 に動画内のあるフレームにおける動画中のキーフレームの内，いくつかのフレーム中における本手法で得られる死角と LiDAR センサから得られる死角の比較を示す．いずれの図においても，上の段から入力画像に本手法で推定された死角を重ね合わせたもの，LiDAR センサから求められる死角領域，本手法で求められる死角領域，それらの和集合を表している．表 1 にそれぞれの画像の IoU，Precision，Recall を示す．LiDAR センサから計算される死角は本手法で得られる死角よりも全体の面積は大きい場合が多い．しかし，本手法で得られる死角は概ね LiDAR から得られる死角の領域内に含まれており，実際 Precision

図 5 あるフレームとその 5 秒後のフレームの LiDAR 点をそれぞれ水色，黄緑色で投影する．壁や停止中の車など固定された物体上に点が重なっていることが分かる．このことから，本手法で求められるカメラの外部パラメータは 5 秒間の間では正しく推定できる．

画像	1	2	3	4	5	6	動画全体
IoU	0.1772	0.2859	0.1251	0.3394	0.2626	0.5660	0.1648
Prec.	0.8996	0.8164	0.8894	0.8234	0.6831	0.6062	0.5785
Rec.	0.1808	0.3056	0.1271	0.3661	0.2991	0.8951	0.2047

表 1 KITTI データセットに含まれるある動画について、真値を LiDAR による死角、推定結果を提案手法による死角としたときの図 7 のフレームと KITTI データセットのある動画全体の IoU, Precision，Recall の値をそれぞれ示す．本手法で得られる死角は真の死角の一部であり，実際高い Precision となっている．

も高い．このことから，本手法で得られる死角は真の死角の一部であると言える．

4.5 失敗例

各データセットの動画の中には，信号待ちのときに撮影され車が動かない動画，夜間に撮影された動画，道路全体に雪が積もっている動画などがある．これらの動画は自己位置推定，もしくはセマンティックセグメンテーションによる道路領域の推定を正しく行うことができないため，死角は生成できなかった．また自己位置推定を行う過程で，特徴点をうまく見つけることができないために大きく経路がずれるなどがあった場合も死角が生成できなかった．

5. 結論

飛び出しの潜在的な危険が潜む死角を画像や動画から求めることは，自動運転の発展に不可欠である．しかし，見えない箇所をアノテーションすることは出来ないため，これまでの手法では死角を直接求めることは難しかった．そこで本研究では，視点移動による景色の変化に着目することで，二次元での死角を定義し，その自動的な計算手法を提案した．

求めた死角を検証するため二つの評価実験を行なった．Visual SLAM で求められたカメラの外部パラメータを用いて，あるフレームとその 5 秒後のフレームの LiDAR 点を世界座標系でプロットした結果，5 秒間のカメラの外部パラメータを正しく求められることが示された．また，LiDAR 点から作成される死角との比較では，本手法で求められる死角が真の死角の一部であることが示された．本手法で計算できる手法は実際に飛び出しの危険がある死角であり，安全な自動運転の実現に遮蔽された領域の認識という観点から寄与した．本手法はカメラパラメータが既知の任意の車載動画に対して適用できるため，データセットは今後も拡大させることが可能である．

本手法で求められる死角はカメラの視野角に大きく依存する．実際，4.3 節で述べたように推定結果に対する真値の Recall は小さいと言える．今後の課題は，観測可能な死角の範囲を増やし LiDAR 点との比較における Recall を向上させることである．そのためには，パノラマ画像を用いる必要があると考えられる．

また，他の課題として，自己位置推定が適用できる動画をさらに増やすため，車載動画中の移動物体のみにマスクをかける処理が必要があると考えられる．Visual SLAM では画像内の特徴点から三次元座標を環境マップに登録し，その三次元座標を用いて外部パラメータの推定を行う．求められた特徴点が移動物体上の点であればその分自己位置推定の精度も低下すると考えられる．

謝辞

この研究の一部は JSPS 20H05951, 21H04893, JST JP-

図 6 [左上] 元画像に対し LiDAR で得られる死角を投影したもの．LiDAR で得られる死角は疎である．[左下] 提案手法により得られる死角．本手法で得られる死角は密である．[右]LiDAR 点群全体 (水色)，LiDAR で得られる死角 (黄色)，本手法で得られる死角 (赤色) を三次元的に投影したもの．緑色のバツ印は視点の位置を表す．

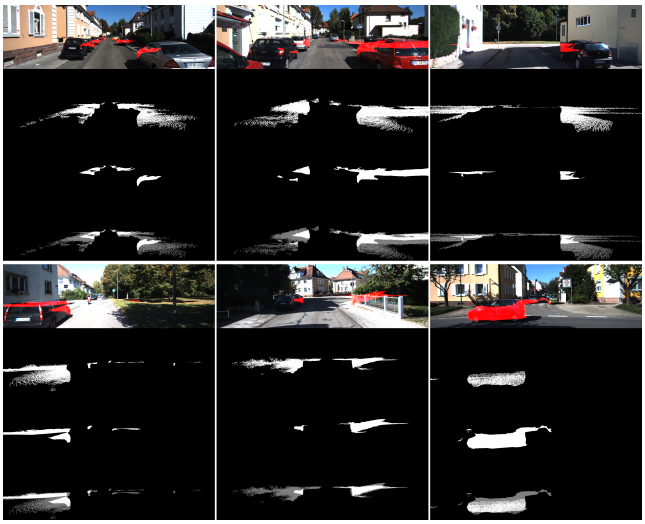

図 7 KITTI データセットに含まれる画像での死角の比較．各画像において上から順に元画像，LiDAR で得られる死角，本手法で得られる死角，それらの和集合を表している．本手法で得られた死角は LiDAR で得られる死角の一部になっている．

MJCR20G7 の助成を受けて行ったものです．

参考文献

[1] Geiger, A., Lenz, P., Stiller, C. and Urtasun, R.: Vision meets Robotics: The KITTI Dataset, *International Journal of Robotics Research* (2013).
[2] Yu, F., Chen, H., Wang, X., Xian, W., Chen, Y., Liu, F., Madhavan, V. and Darrell, T.: BDD100K: A Diverse Driving Dataset for Heterogeneous Multitask Learning, *CVPR* (2020).
[3] Malla, S., Dariush, B. and Choi, C.: TITAN: Future Forecast using Action Priors, *CVPR* (2020).
[4] Barnum, P., Sheikh, Y., Datta, A. and Kanade, T.: Dynamic seethroughs: Synthesizing hidden views of moving objects, *ISMAR*, pp. 111–114 (online), DOI: 10.1109/ISMAR.2009.5336483 (2009).
[5] Zhu, X., Yin, Z., Shi, J., Li, H. and Lin, D.: Generative Adversarial Frontal View to Bird View Synthesis, arXiv:1808.00327 (2018).
[6] Mani, K., Daga, S., Garg, S., Shankar, N. S., Jatavallabhula, K. M. and Krishna, K. M.: MonoLayout: Amodal scene layout from a single image, *CoRR*, Vol. abs/2002.08394 (online), available from ⟨https://arxiv.org/abs/2002.08394⟩ (2020).
[7] Yang, W., Li, Q., Liu, W., Yu, Y., Ma, Y., He, S. and Pan, J.: Projecting Your View Attentively: Monocular Road Scene Layout Estimation via Cross-View Transformation, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15536–15545 (2021).
[8] Watson, J., Firman, M., Monszpart, A. and Brostow, G. J.: Footprints and Free Space From a Single Color Image, *CVPR* (2020).
[9] Sumikura, S., Shibuya, M. and Sakurada, K.: OpenVSLAM: A Versatile Visual SLAM Framework, *ACM MM*, pp. 2292–2295 (2019).
[10] Lasinger, K., Ranftl, R., Schindler, K. and Koltun, V.: Towards Robust Monocular Depth Estimation: Mixing Datasets for Zero-Shot Cross-Dataset Transfer, arXiv:1907.01341 (2019).
[11] Wu, Y., Kirillov, A., Massa, F., Lo, W.-Y. and Girshick, R.: Detectron2, `https://github.com/facebookresearch/detectron2` (2019).