VLM と二輪自動車ドライブレコーダー映像による道路上の運転危険箇所の検出の検討

Detecting Hazardous Road Areas Using Vision-Language Models and Motorcycle Dashcam Videos

片岡 美樹 *，湯村 翼 *
Miki KATAOKA*, Tsubasa YUMURA*
* 北海道情報大学　Hokkaido Informaton University

Keywords : Vision-Language Model，画像処理，二輪自動車

1. は　じ　め　に

二輪自動車は通勤や物流など幅広い用途で利用される．二輪自動車は，四輪自動車と比較して転倒や事故のリスクが高い．その要因の一つとして，接地面積が小さく車体が不安定であることが挙げられる．そのため，同じの路面状況であっても二輪自動車では危険度が大きく異なる．

実際の走行においては，マンホールの蓋，白線，橋の継ぎ目，砂利や砂など，二輪自動車特有の危険に感じやすい路面要素が存在する．しかし，これらは天候や運転者のスキルなどさまざまな環境要因に依存するため，画像処理の物体検出という単純な手法では不十分である．

近年，画像や映像と言語を同時に扱う Vision-Language Model（VLM）が登場し，自然言語によるプロンプトを用いて映像内容を解釈することが可能となった．本研究では，VLM を用いることで，二輪自動車の走行における危険箇所を検出する（Fig. 1）．

ドライブレコーダー映像を用いた危険地点抽出に関する研究は，主に四輪自動車を対象として行われてきた．例えば三浦ら[(1)]は，交差点や出会い頭事故を対象とし，ドライブレコーダーデータから危険地点や危険状況を抽出する手法を検討している．しかし，これらの研究では，二輪自動車特有の路面危険要素は十分に扱われていない．

また，VLM に関する研究では，プロンプト設計が出力結果に大きな影響を与えることが指摘されている．Miyata[(2)] は，VLM エージェントに役割を与えることで評価の質が向上する可能性を示しており，適切な指示設計の重要性が示唆されている．

二輪自動車における路面危険要素の評価には，以下の課題が存在する．

- 走行者の感覚（危険や不安）を含めた評価が必要である

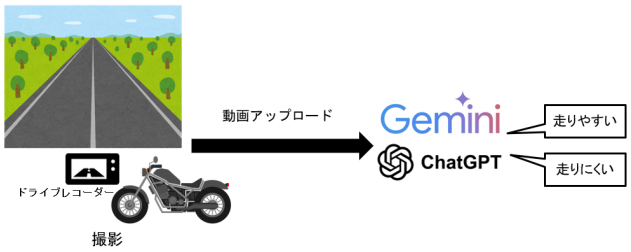


Fig. 1　本研究の概要.

- 複数の路面要素や走行環境を統合的に理解する必要がある

これらの課題を踏まえ，本研究では危険箇所検出において VLM を活用するためのプロンプト設計に着目する．本研究の目的は，マンホールの蓋，白線，砂利などの二輪自動車特有の路面危険要素に着目し，VLM による危険箇所検出の可能性を検討することである．特に，プロンプトの与え方による出力の違いを比較し，どのような条件で有効な検出が可能となるかを明らかにする．

2. 実　験　概　要

本研究では，二輪自動車の走行に適した道路の存在する危険箇所を検出できる可能性を検討するため，ドライブレコーダー映像に対して VLM を適用する実験を行った．二輪自動車から撮影した走行映像を VLM に入力し，道路環境に含まれる危険要素を言語的に抽出させることで有効性を確認する．

2·1　使用した VLM と実験設計

本実験では，画像・映像と自然言語を同時に扱うことが可能な VLM として，Gemini 3 Flash および ChatGPT 5.2 を使用した．両者に対して同一の映像とプロンプトを入力し，出力結果を比較することで，二輪自動車の危



Fig. 2　実験 1〜3 で用いた動画のスナップショット．

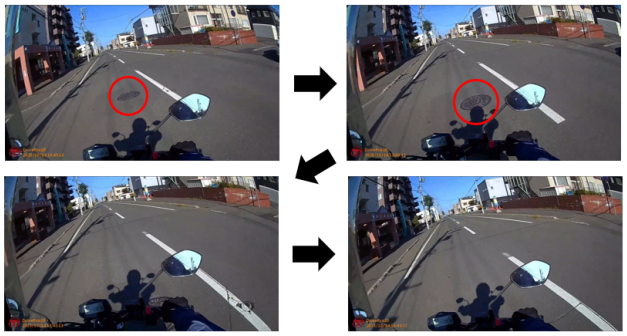


Fig. 3　実験 4 で用いた動画のスナップショット．2 つのマンホールの蓋が映る．

険箇所検知への適用可能性を検討した．

入力映像は，二輪自動車に搭載したドライブレコーダーによって撮影した走行映像である．

ドライブレコーダーの映像は，北海道江別市および札幌市を自動二輪車で走行して収集した．走行車両として SUZUKI GIXXER 150 を，ドライブレコーダーとして Transcend DrivePro 20 を使用した．撮影した動画は，累計動画長が 240 分，累計容量は 38.8GB であった．なお，ドライブレコーダーの仕様で，1 本あたりの動画は最大 5 分で分割される．

3. 実験内容と結果

3·1　概要

撮影した動画を用い，様々なプロンプトを用いて実験を実施した．本論文では，その中で特徴的な 4 つの実験について記述する．実験 1〜3 は Gemini 3 Flash を，実験 4 は Gemini 3 Flash と ChatGPT 5.2 を VLM として用いた．実験 1〜3 では，札幌市内で撮影した 5 分間の同一の動画（Fig. 2）を用いた．

3·2　実験 1：単純なプロンプトによる検証

実験 1 では，単純なプロンプトを用いて，バイクの危険箇所の判別可能性を検証した．実験 1 に用いたプロンプトと，その結果を Table 1 に示す．単純なプロンプトでは，抽象的な結果しか得られない．そのため，自動二輪車における危険箇所を具体的に指示する必要があることが明らかとなった．

3·3　実験 2：危険箇所の列挙

実験 2 では，自動二輪車における危険箇所の候補をプロンプトによって与えることで，より具体的な回答の取得可否を検証した．危険箇所の候補は，自動二輪車での走行経験から，白線，工事用の鉄板，橋の継ぎ目，砂利・砂・泥，ひび割れ，大きな穴，段差，アスファルトの波打ち，路面の縦溝，枯葉・枯れ枝・苔，見通しの悪いカーブ，交通量の多い市街地，トンネル・橋の上・ビルの谷間山道・林道とした．実験 2 に用いたプロンプトと，その結果を Table 2 に示す．危険箇所の回答は，動画のフレーム番号で示された．この結果は，動画のフレームを分析して照合すれば有用であるものの，そのまま人間が用いることはできない．

3·4　実験 3：時刻での回答取得

実験 3 では，実験 2 と同様に危険箇所候補をプロンプトで指示した上で，動画内の時刻による回答の取得可否を検証した．実験 3 に用いたプロンプトと，その結果を Table 3 に示す．目的通り，動画内の時刻で回答を得ることができた．

3·5　実験 4：モデルの精度比較

VLM のモデルによる回答の差異を検証するため，Gemini 3 Flash と ChatGPT 5.2 に対して全く同じ動画とプロンプトを用いて回答を比較した．危険要素をマンホールの蓋に限定し，映像長を 5 秒に編集した短時間動画 (Fig. 3) を用いた実験を行った。プロンプトでは，動画内に映る 2 つのマンホールを検出し，その時刻を掲示するという単純なタスクを与えた．実験 4 に用いたプロンプトと，その結果を Table 4 に示す．ChatGPT 5.2 では，2 つのマンホールを正しく検出し，動画内での映り方についても詳しく回答した．一方，Gemini 3 Flash では，2 つしかないマンホールを，1 度の誤検出を含む 3 度検出した．追加のプロンプトで，誤検出の原因を探ったが，特定することはできなかった．同一の動画及びプロンプトでも，モデルによる差異が生じることが確認できた．

4. 結　　果

本節では，VLM を用いた道路危険検出実験の結果について整理する．本研究では定量的な性能評価は行わず，VLM の出力内容や傾向を比較し，観察結果をもとに評価した．

複数の危険要素を同時に指定し，5 分間のドライブレコーダー映像を入力した実験では，危険箇所はフレーム番号で提示された．そのため，映像上の具体的な位置を把握することが困難であった．次に，危険箇所を時間情報で出力するようプロンプトを変更した．この条件では，映像の冒頭については評価が行われていたものの，その後の区間では十分な出力が得られなかった．

危険要素をマンホールの蓋に限定し，5 秒間の短時間動画を入力した実験では，VLM として Gemini を使用した．この条件では，マンホールの蓋が存在しない箇所を存在すると判断する例や，実際に存在する箇所を検出できない例が確認された．

同一の映像およびプロンプト条件のまま，VLM を Gemini から ChatGPT に変更した実験では，マンホールの蓋を適切に判断できた．危険箇所として示された時間や位置も映像内容と整合していた．

以上の結果から，映像長や危険要素の指定方法，使用する VLM の違いによって出力結果に差が生じることが確認された．一方で，評価基準の明確化や定量的な検証は行えていないため，結果の妥当性については今後の検討が必要である．

5. お　わ　り　に

本研究では，二輪自動車の走行に適した道路を検出する可能性を検討するため，ドライブレコーダー映像に VLM を適用し，道路上の危険要素を検出する実験を行った．二輪自動車は路面状態の影響を受けやすい一方，既存の危険検知手法の多くは四輪自動車を前提としている．本研究は，VLM を用いて二輪自動車特有の危険要素を捉える試みである．実験の結果，映像長やプロンプト内容，使用する VLM の違いによって出力結果に差が生じることが確認された．特に，短時間映像において ChatGPT を用いた場合には，マンホールの蓋を適切に判断できる可能性が示された．これにより，VLM が二輪自動車向けの道路危険検出に応用可能であることが確認できた．一方で，本研究では定量的な評価は行えておらず，結果の妥当性は主観的判断に基づいている．今後は，評価基準の明確化や客観的指標の導入，多様な走行環境を対象とした検証が課題である．

文　　献

(1) 三浦紘司，廣畠康裕，松尾幸二郎ドライブレコーダーデータを用いた交通事故危険地点・危険状況の抽出に関する検討，土木計画学研究・講演集， Vol. 38, p. 289 (2008).

(2)Miyata, A.: Can Role-Based LLM Dialogue Improve Understanding of Accessibility Issues?, *Proceedings of the 27th International ACM SIGACCESS Conference on Computers and Accessibility*, pp. 1–5 (2025).

Table 1　実験 1 のプロンプトと結果.

プロンプト	回答
この道はバイクにとって走りやすい道ですか？	1. 路面状況 舗装の荒れ: 映像の序盤から、ひび割れやパッチワーク（補修跡）が目立ちます。特にコーナリング中にこうした段差に乗り上げると、フロントが跳ねたりタイヤの接地感が失われやすいため、注意が必要です。浮き砂・落ち葉: 道幅が狭く、山道特有の砂利や枯れ葉が路肩側に溜まっています。ライン取りをミスして端に寄るとスリップの危険があります。 2. 道幅とカーブ ブラインドコーナー: 先の見通しが効かないタイトなカーブが連続しています。対向車（特に軽トラックや作業車）がセンターラインを越えてくる可能性を常に想定しておく必要があります。狭い離合ポイント: バイク同士なら問題ありませんが、車と遭遇した際はすれ違いに気を使う幅です。 3. メリット 絶景: 途中で見える開けた景色は素晴らしく、ゆっくり流す分には非常にリフレッシュできるルートです。交通量: 映像を見る限り、主要幹線道路に比べて車通りが少なそうです。自分のペース（控えめな速度）を保ちやすいのはメリットと言えます。

Table 2　実験 2 のプロンプトと結果.

プロンプト	回答
この動画のどこにマンホールの蓋、白線、工事用の鉄板、橋の継ぎ目、砂利・砂・泥、ひび割れ、大きな穴、段差、アスファルトの波打ち、路面の縦溝、枯葉・枯れ枝・苔、見通しの悪いカーブ、交通量の多い市街地、トンネル・橋の上・ビルの谷間山道・林道といった危険性がありますか？	マンホールの蓋 [Image 59], [Image 72], [Image 94], [Image 169] 白線 [Image 2], [Image 32], [Image 55], [Image 170] 工事用の鉄板 [Image 133] 橋の継ぎ目 [Image 55] , [Image 140], [Image 178] 砂利・砂 [Image 148] , [Image 161] 枯葉・枯れ枝 [Image 161], [Image 168]

Table 3　実験 3 のプロンプトと結果.

プロンプト	回答
この動画の何分何秒にマンホールの蓋、白線、工事用の鉄板、橋の継ぎ目、砂利・砂・泥、ひび割れ、大きな穴、段差、アスファルトの波打ち、路面の縦溝、枯葉・枯れ枝・苔、見通しの悪いカーブ、交通量の多い市街地、トンネル・橋の上・ビルの谷間山道・林道といった危険性がありますか？	0:02 - 0:03 マンホールの蓋 0:05 - 0:07 砂利/砂 0:13 - 0:14 工事用の鉄板 0:14 - 0:15 白線 (横断歩道) 0:21 - 0:22 橋の継ぎ目 0:31 - 0:33 アスファルトの波打ち

Table 4　実験 4 のプロンプトと結果.

プロンプト	回答
[Gemini 3 Flash] 側溝や小さな点検口を除いて、マンホールの蓋はこの動画の何秒の部分にありますか？バイクがマンホールの蓋の場所を通過する秒数を表示してください。	00:01 付近（走行車線の中央付近） 00:03 付近（走行車線の中央付近） 00:04 付近（走行車線の中央付近）
[ChatGPT-5.2] 側溝や小さな点検口を除いて、マンホールの蓋はこの動画の何秒の部分にありますか？バイクがマンホールの蓋の場所を通過する秒数を表示してください。	1. 約 1.2 秒付近 1.0 秒頃に前方に現れ、1.17〜1.27 秒あたりでバイク直下を通過しています。 2. 約 4.25 秒付近 4.0 秒頃に前方に現れ、4.2〜4.3 秒あたりでバイク直下を通過しています。