

イベントカメラと動的ライン照明を用いた 複数の半透明レイヤの深度推定

木渕 寛太¹ 杉本 志織² 曾我部 陽光² 久保 尋之¹

概要：本研究では、イベントカメラとライン照明を用いて撮影したイベントの時間情報から、シーンに複数置かれた半透明なレイヤ状の被写体の深度を推定する方法を提案する。具体的には、カメラから被写体までの深度に応じて照明が当たる時刻が異なるように、ライン照明を高速に繰り返し動かしながらシーンに照射する。これを高い時間分解能で受光強度変化を記録可能なイベントカメラで観測することにより、イベントの発生時刻から照明座標系での照射位置を特定し、カメラ座標系での観測位置と対応付けることにより被写体の深度を推定する。このとき、カメラから見て深度の異なる複数のレイヤが同じ位置に重なっているような場合には、カメラの各画素では各レイヤからの反射によるイベントが混在して観測される。そこで照明の照射周期が既知であることを利用し、画素毎に観測された混合イベント列と照射周期に基づいて生成したテンプレート信号の相互相関関数から、各レイヤ由来の周期的なイベント列の位相をそれぞれ推定することで、各レイヤへの照明の照射時刻を特定し深度推定を実現する。実験により、異なる深度にある2つの半透明なレイヤそれぞれの深度が推定できていることを確かめた。

キーワード：イベントカメラ、アクティブステレオ法、深度推定

1. はじめに

近年のマシンビジョン研究の発展はめざましく、例えば自動車の自動運転技術などの開発が盛んに行われており、それに伴ってシーンの立体的な構成を推定する技術が注目されてきている。シーンの立体的な構成、すなわち深度の情報は、シーンに含まれる物体の位置を表すため、前述した自動運転技術に限らず、ファクトリオートメーションなどで幅広く活用されている。

シーンの深度を推定する手法としては、これまでカメラを用いて非接触で推定する様々な手法が研究されてきた。カメラによる深度推定手法には大きく分けて、2つのカメラを用いて三角測量の原理で深度推定を行うパッシブステレオ法 [5], [6] と、パッシブステレオ法の片方のカメラを光を投影する機器に置き換えたアクティブステレオ法 [3], [8] の2つがある。パッシブステレオ法では、異なる位置に置かれた複数のカメラでシーンを撮影し、得られた複数の画像から対応する点を検出する。さらに、三角測量 [9] の原理でその点の深度推定を行う。アクティブステレオ法で

は、まずプロジェクタなどの投影器を用いてスポット光やスリット光などを投影する。この様子をカメラで撮影して得られた反射光を手がかりに、カメラとプロジェクタの対応する画素を特定し、同様に三角測量の原理で深度推定を行う。

この手法では、被写体が半透明で光の一部を透過し一部を反射するレイヤを複数備えるとき、プロジェクタの別々の画素に対応する照射光が各レイヤを透過・反射した結果カメラの同じ画素で混在して観測されることにより、カメラの各画素に対応するプロジェクタの画素を特定できないという問題がある。

そこで本研究では、アクティブステレオ法のカメラをイベントカメラ [2] と呼ばれる特殊なカメラに置き換えることで複数レイヤの深度をそれぞれ推定する手法を提案する。一般的なカメラが各画素に入射する光の強さの積算値である輝度値をフレームごとに同期して記録するのに対し、イベントカメラは“イベント”とよばれる各画素に入射する光の強さの微分値である輝度変化量と変化が起きた時刻を画素毎に非同期で記録する [10]。イベントカメラではマイクロ秒単位の高い時間分解能でイベントを観測することができるため、プロジェクタの画素毎に少しずつ照射時刻をずらすことで、それぞれの照射光に由来するイベントを独立に観測することが可能になる。従来のイベントカメラと

¹ 東海大学 情報通信学部 情報メディア学科
〒108-8619 東京都港区高輪 2-3-23

² NTT コンピュータ&データサイエンス研究所
〒239-0847 神奈川県横須賀市光の丘 1-1

アクティブ照明による研究 [4] では、単一のレイヤのみを想定していたが、本研究では、後述の通り複数のレイヤが含まれることを想定した処理方法を導入している。カメラのある画素で観測される複数のイベントについて、各イベントをレイヤに対応付け、イベント観測時刻からプロジェクタの対応画素を特定することで、その画素で観測される複数のレイヤの深度を推定することができる。

以降、2章では本手法の詳細について説明し、3章では実際の計測と推定を実施して本手法の有効性を検証する。4章ではまとめと今後の課題を述べる。

2. 提案手法

本章では、まず初めにイベントカメラの性質を簡単に紹介し、イベントカメラを用いたアクティブステレオ法の動作原理について概説する。さらに、本論文で提案する、複数のレイヤ由来のイベントが混在した混合イベント列をレイヤごとの独立イベント列に分離して深度推定を行う方法について詳細に説明する。

2.1 イベントカメラとは

一般的なカメラでは、全画素の輝度値をある決められた時刻のフレームごとに記録するのにに対し、イベントカメラでは、イベントと呼ばれる輝度の時間変化量をセンサの各画素で非同期に記録することが可能である。イベントカメラの各画素で記録されるイベントデータには、イベントが発生した時刻 t 、画素の位置 (x, y) に加え、時間の経過により輝度が大きくなった時に観測されるポジティブイベントと、時間の経過により輝度が小さくなった時に観測されるネガティブイベントとを区別するためのフラグが含まれる。なお、イベントカメラにはマイクロ秒単位の高い時間分解能でイベントを記録できるという特徴がある。本研究では、イベントカメラを使って計測を行った際に画素 (x, y) において n_+ 回のポジティブイベントが発生するとき、全てのポジティブイベントの発生時刻を要素とする集合を

$$E_+(x, y) = \{t_{+1}, t_{+2}, \dots, t_{+n_+}\} \quad (1)$$

と表記する。ただし t_{+i} は画素 (x, y) で観測された i 番目のポジティブイベントの発生時刻である。

2.2 イベントカメラを用いたアクティブステレオ法

単純なアクティブステレオ法では、プロジェクタからスポット光やスリット光などを投影し、カメラで撮影した反射光からプロジェクタの対応画素を特定して深度推定を行っている [8]。ここで図 1 のような半透明のレイヤが複数重なるシーンでは、輝度値を記録する一般的なカメラでは複数のレイヤからの光が混在して観測されてしまい対応画素の特定ができない (図 1(a))。本研究ではカメラを各画素で非同期的に輝度の変化量を高時間分解能で記録する

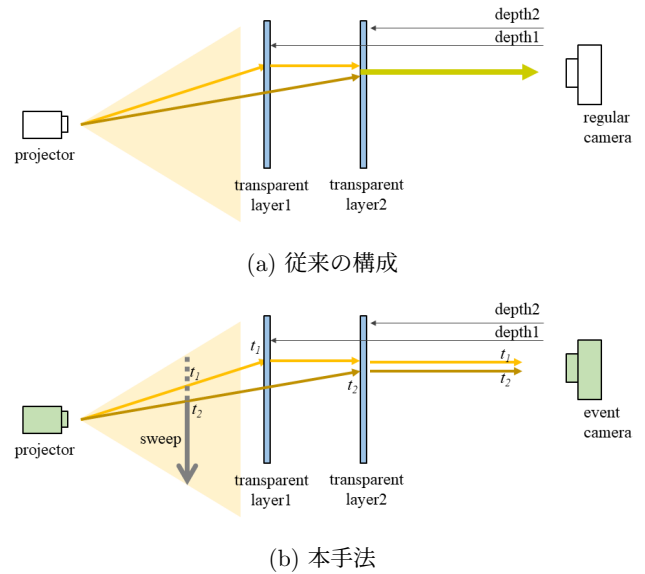


図 1: 従来のアクティブステレオシステムと本研究の計測システム。(a) 従来のカメラで透明なレイヤが2層あるシーンを観測する場合、輝度値の情報だけでは奥のレイヤ (layer1) からの光と手前のレイヤ (layer2) からの光とを区別することが出来ない。(b) プロジェクタでライン照明をスイープしながら照射し、さらにそれぞれのレイヤに当たった光が散乱する時刻を高精度に記録することで複数レイヤの深度値の推定を可能とする。

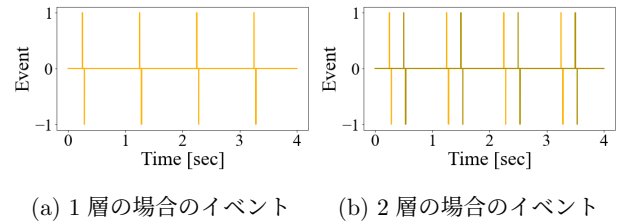


図 2: シーンに含まれるレイヤが1層の場合 (a) と2層の場合 (b) における記録されるイベントの比較。1層の場合には照明の1周期あたり1回のイベントが発生するが、2層の場合には2回のイベントが発生する。

イベントカメラに置き換え、観測したイベントをレイヤごとに分離して対応画素の特定を行う (図 1(b))。

まずプロジェクタとカメラは平行化されているものとし、光切断法 [7] と同様にプロジェクタでライン状の光線 (ライン照明) をシーンに照射する。このとき、ライン照明は少しずつ方向を変えながらスイープすることとし、端に達した時点で再度もう一方の端から照射を再開する。周期的に照射を行うこととすると、ライン照明の照射位置はこの周期に対する位相に対応する。従って、シーンからの散乱光をイベントカメラで観測することで、各画素で散乱光が捉えられた時刻に基づいて照射周期に対する位相を求め、ライン照明の照射位置を導きだし、三角測量の原理で当該画素の深度情報を得ることが可能となる。

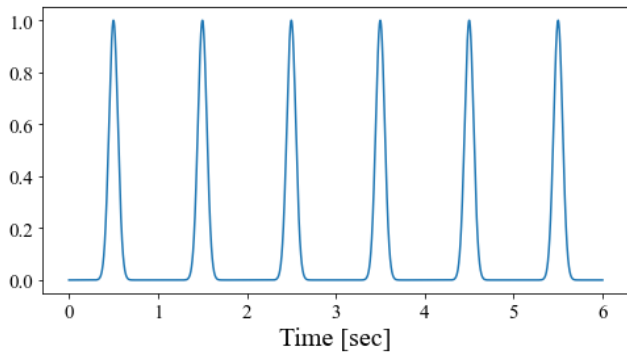


図 3: 観測されるイベント情報をモデル化したテンプレート信号. 本研究ではガウス関数図 2 を使用した.

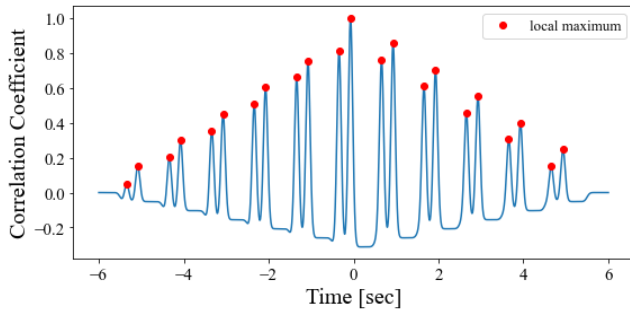


図 4: 図 3 のテンプレート信号とイベント信号の相互相関関数. 本図の作図に使用したイベント信号は複数レイヤを撮影したときに記録された実際のデータを使用している.

図 2(a) に単一のレイヤに照明を照射したときに観測されるイベントを示す. 照明が照射されると輝度が大きくなりポジティブイベントが記録され, その後照射が終わると輝度が小さくなりネガティブイベントが記録される. 一方, 図 2(b) に複数の半透明なレイヤに照明を照射したときに観測されるイベントを示す. イベントは各レイヤで輝度の変化が生じる毎にマイクロ秒単位で逐次記録されるため, 同じ画素に複数のレイヤからの反射光が入射する場合, 異なる位相で同一周期のイベント信号を複数足し合わされたような混合イベント列が記録される. 各イベント信号の周期は照明の照射周期に一致し, 位相は照明の照射位相に対応する. この混合イベント列を各レイヤに対応する独立イベント列に分離し, 各独立イベント列の位相を推定することで, 各レイヤの深度推定を行うことができる. 本論文では位相推定までを行い, 深度推定は行わない.

2.3 位相推定

位相推定は, 画素毎にイベント信号とテンプレート信号の相互相関関数を計算し, 得られた相互相関関数について極大値をとる位相を求めることによって行う.

テンプレート信号は図 3 に示す照明の照射周期 T と同じ周期を持つ信号であり, 単一のレイヤ由来で発生する独立イベント列をモデル化したものである. このテンプレ

ート信号と混合イベント信号の相互相関関数を計算することで, 各レイヤに対応する独立イベント信号を分離し, テンプレート信号に対する各独立イベント信号の位相差を求めることができる. 図 4 にテンプレート信号とイベント信号との相互相関関数を示す. 2 つの類似信号の間に位相差がある場合, 相互相関関数は原点から位相差分ずれた時刻に極大値を持つ. つまり相互相関関数が極大となる時刻を $-T/2$ から $T/2$ の範囲から特定することで, テンプレート信号と各独立イベント信号の間の位相差をそれぞれ求めることができる.

テンプレート信号は独立イベント列をモデル化したものであるため, 理想的には周期的に繰り返されるパルス信号の列である. ただし, 実際のイベントカメラは輝度の変化量が一定のしきい値を超えた時刻にイベントを記録するが, センサの性質によっては光量や帯域の不足などの様々な原因により, 輝度の変化が起きたタイミングから遅れてイベントが記録されることがある. この遅延による周期のずれを吸収するため, テンプレート信号の各周期のパルスに幅を持たせる必要がある. 従って, 本論文では, テンプレート信号を周期的に繰り返されるガウス信号としてモデル化した. また, 一度に多量のイベントが発生するなどして帯域が不足した場合に輝度変化がイベントとして検出されないイベントドロップが起きることもある. このため, テンプレート信号の長さを複数周期分にするすることで, 一時的なイベントドロップの影響を軽減した.

ガウス関数を用いたテンプレート信号を $\phi(t)$ とすると,

$$\phi(t) = \sum_{i=1}^n \exp\left(-\frac{(t-T_i)^2}{2\sigma^2}\right) \quad (2)$$

で表される. なお, σ は関数の広がりを決める定数である. 図 3 に式 (2) を用いて作成したテンプレート信号の例を示す. このテンプレート信号とイベント信号の相互相関関数を式 (4) によって計算する.

図 4 に計算された相互相関関数の例を示す. なお, 照明が照射されるとポジティブイベントとネガティブイベントは対となって記録されるため, 計算に使うイベント信号はポジティブイベントデータのみから生成する (式 (3)).

$$e_+(x, y, t) = \begin{cases} 1 & \text{if } t \in E_+ \\ 0 & \text{otherwise,} \end{cases} \quad (3)$$

$$c(x, y, t) = \{\phi(x, y, t) * \hat{e}_+(x, y, t)\}(t) \quad (4)$$

$\hat{e}_+(x, y, t)$ は $e_+(x, y, t)$ を時刻 t に沿って正規化したものである. なお, $*$ は畳み込み演算子である.

また, 相互相関関数の 1 周期内に現れる極大値の個数は, レイヤの枚数に一致すると考えられる. そのため, 極大値の個数に応じてその画素におけるレイヤの枚数と各レイヤに対応する位相を特定できる. 従って, 任意の N 層のレイ

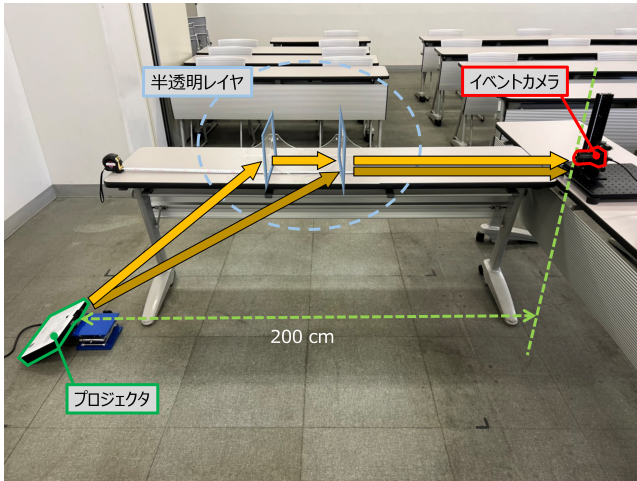


図 5: 撮影環境. 画像は複数レイヤでの撮影の様子. 単一レイヤの撮影では半透明レイヤを 1 枚にして行う. なおこの図の撮影のため手前の机を除いてあるが, 実際の計測では手前にも机が設置されている.

ヤに対して位相を推定することができると考えられるが, 実際には極大値はノイズの影響によりレイヤ枚数以上の極大値が現れることがあるため, 極大値に対する閾値処理を行ってこれを除去する.

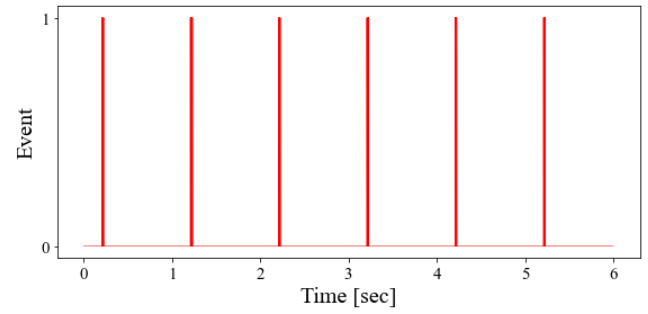
3. 実験

実際のシーンに半透明なレイヤを設置し, 前章で述べた手法を用いて位相を推定が出来ることを実験により確かめる. なお, 本研究では被写体の具体的な深度を求めることはせず, 推定された位相について可視化して評価する.

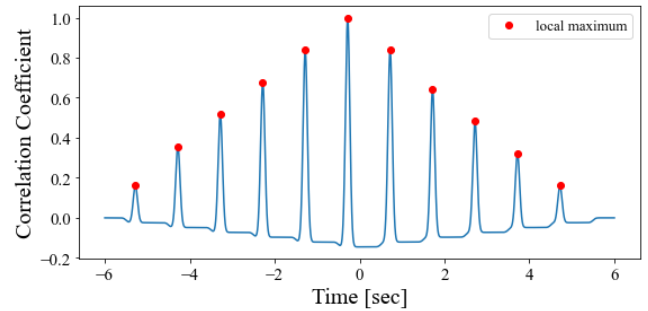
3.1 撮影環境

図 5 に実際の撮影環境を示す. 被写体に対して平行にカメラを設置し, プロジェクタは被写体を挟んでカメラと反対側, 被写体の斜め下にカメラと垂直面を共有するように設置し, 被写体に対して斜め上方向から斜め下方向に向かってライン照明をスイープ照射することで, 被写体の奥行毎に照明が当たる時刻が異なるようにした.

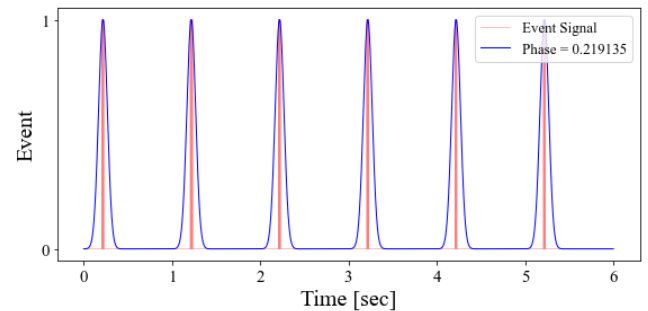
イベントカメラは, Prophesee 社の EVALUATION KIT - Gen4 HD を使用した. このカメラの解像度は 1280×720 であり, 最大の時間分解能は $1 \mu\text{s}$ である. プロジェクタは, 照射周波数 60fps の EPSON 社の液晶プロジェクタ EB-1795F を使用する. このプロジェクタの投影方式は 3LCD 方式であり, 走査周波数は水平 75kHz, 垂直 60Hz である. また, パネル画素数は 1920×1080 である. 被写体である半透明レイヤとして, リアプロジェクションフィルムを貼ったアクリル板を使用する. なお, 一度に多量のイベントが発生してイベントドロップが起きることを防ぐため, リアプロジェクションフィルムはアクリル板の一部のみに貼付し, イベントが一度に発生する量を抑制する.



(a) 混合イベント信号



(b) 相互相関関数



(c) フィッティング結果

図 6: 単一レイヤでの位相推定結果. (a) は画素 $x = 640$, $y = 280$ でのイベント信号 e_+ である. (b) は (a) とテンプレート信号との相互相関関数である. (c) は (a) と推定された位相分ずらしたテンプレート信号とのフィッティング結果である. 凡例に推定された位相の値を示している.

計測は, 照明を半透明レイヤに照射しながら, 半透明レイヤで起きた散乱光をイベントカメラで記録することにより行う. 照明は, 図 5 のように半透明レイヤの背後から周期 $T = 1[\text{sec}]$ として水平方向のライン照明をスイープしながら照射する.

3.2 単一レイヤの位相推定

まず, 単一の半透明レイヤについて位相推定が可能であることを実験により確かめた. 半透明のレイヤをカメラから 100cm の距離に 1 枚だけ配置して撮影を行った.

図 6 に $e_+(x = 640, y = 280)$ の混合イベント信号と, テンプレート信号との相互相関関数と, 推定された位相に基

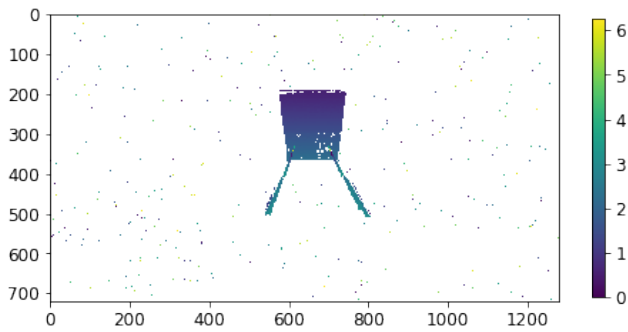


図 7: 単一レイヤの位相推定結果. 位相の単位はラジアン.

づくテンプレート信号と混合イベント信号とのフィッティング結果を示す. また図 7 に各画素で推定された位相を輝度値として可視化したものを示す. 本実験ではライン照明をカメラから見て奥から手前にスイープさせているため, 手前にある被写体ほど大きな位相を持つ. なお, 相関の極大値に対して閾値 0.8 で切り捨てて処理を行ったものについて掲載している.

図 6(c) を見ると, イベント信号に対して正しくフィッティングが行われていることがわかる. 図 7 では, 半透明レイヤが存在する画素に推定結果がプロットされていることが確認できる. また, y 軸方向下の画素になるにつれて, 位相の値が増加していることが確認できる. これはライン照明が上から下の方向にスイープしながら照射されるためであり, 下方向に位相の値が増加するのは正しく位相が推定されていることを示している.

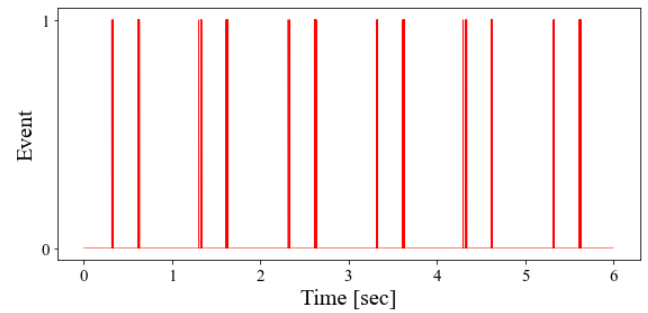
一方, 半透明レイヤが置かれていないはずの画素でも何らかの位相が推定されている画素が散見された. 原因として, アクリル板のリアプロジェクションフィルムが貼られていない部分で起きた散乱光や, スクリーンの下にある机からの反射光が考えられる. 背景部分など直接照明が当たっていない部分でも位相推定されている画素があるが, 原因として熱ノイズ等のセンサノイズに由来するノイズイベント [1] が発生していたことなど考えられる.

3.3 複数レイヤが重なるシーンでの計測

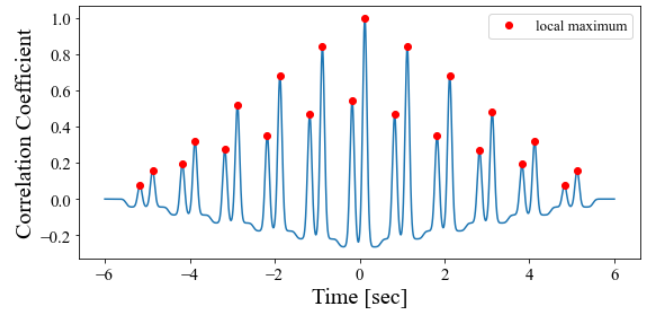
深度の異なる複数の半透明レイヤが重なるとき各レイヤで位相推定が可能か実験により確かめた. 半透明レイヤをカメラから 85cm, 115cm の距離に 2 枚重なるように配置して撮影を行った.

図 8 に $e_+(x=640, y=280)$ のイベント信号および, テンプレート信号との相互相関関数とフィッティング結果を示す. また, 図 9 に位相推定の結果を示す.

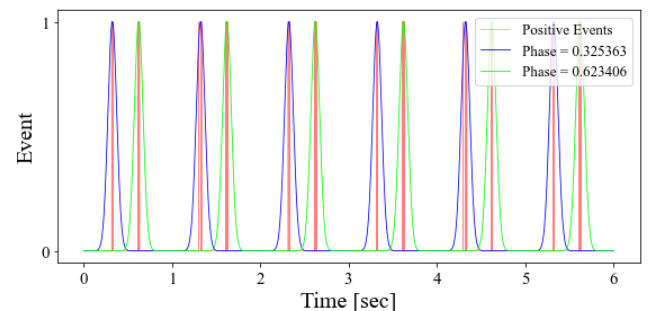
図 8(b) からは, 図 6(b) と比較して周期的にみられる極大値が 2 つずつ増えており, 提案法によって位相を 2 つのレイヤからのイベント信号で分離して, それぞれ推定できることが分かる. 図 8(c) をみると, 2 つのテンプレート信号をイベント信号にそれぞれ正しくフィッティングできて



(a) 混合イベント信号



(b) 相互相関関数



(c) フィッティング結果

図 8: 複数レイヤでの位相推定結果. (a) は画素 $x=640, y=280$ でのイベント信号 e_+ である. (b) は (a) とテンプレート信号との相互相関関数である. (c) は (a) と推定された位相分ずらしたテンプレート信号とのフィッティング結果である. 凡例に推定された位相の値を 2 つ示している.

おり, 混合イベント信号から各独立イベント信号の位相を正しく推定できていることがわかる.

さらに, 図 9(a) に位相推定の結果を, 画像の縦横を xy 軸に, 位相を z 軸とした 3 次元座標上に表示したものを示す. この図によると, 被写体が大きく 2 つの平面状のグループに分かれているおり, どちらのグループでも y 軸方向の下に向かうにつれて位相の値が増加している. これらのことから, 2 つのグループはシーンに設置した各半透明レイヤに対応していると考えられる. 推定結果を位相に対する閾値で 2 つのグループに分け, それぞれ画像として可視化したものを図 9(b) と図 9(c) に示す, 図 9(b) と図 9(c) に示すとおり, レイヤが複数になったとしてもそれぞれの

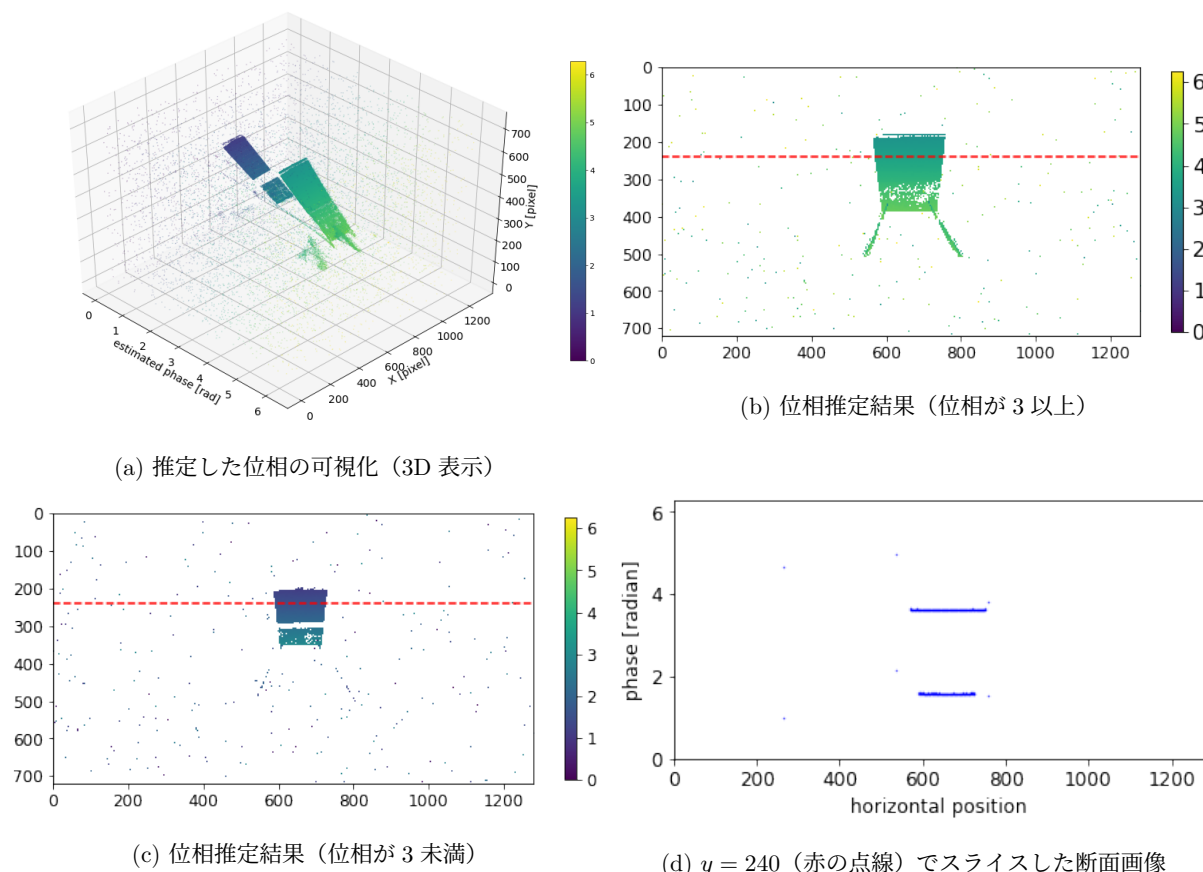


図 9: 複数レイヤが重なるときの深度推定結果

レイヤについて位相推定が行えていることが分かる。

通常のアクティブステレオ法では、プロジェクタとカメラとはそれぞれキャリブレーションを行い、平行化してから深度値を推定することが一般的である。本研究では簡単のためプロジェクタとカメラのキャリブレーションは実施せず、平行化を適用しない状態で位相も推定を実施した。そのため、推定された位相と実際の深度値の関係は厳密には明らかではないが、プロジェクタとカメラの y 軸はできるだけ共通するように設置しており、カメラ上のある y で指定される直線上の画素において、位相が同一であれば深度も同一であるものとした。図 9(d) が示すとおり、レイヤ毎に一定の位相が推定できている事がわかる。

4. まとめと今後の課題

本研究では、プロジェクタ照明の照射時刻のずれを利用して、イベントカメラによる複数の半透明レイヤの深度推定に必要な位相情報の獲得手法を提案した。具体的には、撮影されたイベント信号とテンプレート信号の相互相関関数を計算することで、半透明レイヤの枚数の特定と各レイヤに対応する位相の推定を行った。実験では、半透明レイヤが単一のシーンでレイヤからのイベント信号の位相を推定できていることが確認でき、また複数の半透明レイヤが重なるシーンにおいても各レイヤからのイベント信号の位

相をそれぞれ推定が可能であることを確かめた。複数の半透明レイヤの深度推定が可能になることで、半透明物体の形状復元やガラス越しの物体の深度推定などの手法の開発や応用が進むことが期待される。

今後、プロジェクタとカメラのキャリブレーションを実施し、位相情報を用いて深度情報が復元可能であることを示すことが課題である。また、本研究ではシーンに含まれる最大のレイヤ数は 2 であるものと仮定して実験を行った。基本的には同じ手法でより多くのレイヤ数にも対応可能であるが、層の増加に伴ってカメラで捉えられる散乱光の強度が弱くなっていくと考えられる。さらに、本稿では照明とカメラが対向して設置しており、半透明レイヤで生じる前方散乱光を捉えて深度の推定を行っている。同様の手法で照明とカメラを同じ方向に設置し、半透明レイヤで生じる後方散乱光（ないしは反射光）を捉えて深度の推定することも可能と考えられる。但し、一般に前方散乱光より後方散乱光の方が弱いいため、ここでも光量の不足が問題になると考えられる。イベントカメラのセンサでは、光量が不足することによりイベント検出遅延やノイズが増加することが知られているため、今後は S/N 比の向上や、位相推定アルゴリズムの改良などが課題になると考えられる。

また、本実験ではテンプレート信号としてガウス関数を用いたが、実際のイベントカメラのセンサ特性に応じた、

より尤もらしいテンプレート信号のモデル化を行うことも今後の課題である。

謝辞 本研究の一部は JST 創発的研究支援事業 (PMJFR206I), CREST(JPMJCR1764), 科研費 JP19H04138, JP21K19799 の助成を受けた。

参考文献

- [1] Feng, Y., Lv, H., Liu, H., Zhang, Y., Xiao, Y. and Han, C.: Event Density Based Denoising Method for Dynamic Vision Sensor, *Applied Sciences*, Vol. 10, No. 6 (online), DOI: 10.3390/app10062024 (2020).
- [2] Gallego, G., Delbruck, T., Orchard, G., Bartolozzi, C., Tabatabaei, B., Censi, A., Leutenegger, S., Davison, A., Conradt, J., Daniilidis, K. and Scaramuzza, D.: Event-based Vision: A Survey, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 44, pp. 154–180 (2022).
- [3] Ikeuchi, K. and Horn, B. K.: Numerical shape from shading and occluding boundaries, *Artificial Intelligence*, Vol. 17, No. 1, pp. 141–184 (online), DOI: [https://doi.org/10.1016/0004-3702\(81\)90023-0](https://doi.org/10.1016/0004-3702(81)90023-0) (1981).
- [4] Matsuda, N., Cossairt, O. and Gupta, M.: MC3D: Motion Contrast 3D Scanning, *2015 IEEE International Conference on Computational Photography (ICCP)*, pp. 1–10 (online), DOI: 10.1109/ICCPHOT.2015.7168370 (2015).
- [5] Okutomi, M. and Kanade, T.: A multiple-baseline stereo, *Proceedings. 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 63–69 (online), DOI: 10.1109/CVPR.1991.139662 (1991).
- [6] Tsai, R. Y.: Multiframe Image Point Matching and 3-D Surface Reconstruction, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-5, No. 2, pp. 159–174 (online), DOI: 10.1109/TPAMI.1983.4767368 (1983).
- [7] 新津靖, 宮崎新一: 光切断法を用いた非接触 3 次元全周形状計測システム, 精密工学会誌, Vol. 62, No. 12, pp. 1790–1794 (オンライン), DOI: 10.2493/jjspe.62.1790 (1996).
- [8] 梅田和昇, 寺林賢司: アクティブステレオ法による距離画像センサー, 光学, Vol. 41, No. 5, pp. 275–280 (2012).
- [9] 矢野直樹, 金谷健一: 平面シーンの最適三角測量, 情報処理学会研究報告. CVIM, [コンピュータビジョンとイメージメディア] (2010).
- [10] 國益良太, 岡部孝弘: 偏光イベントベースカメラによる拡散反射と鏡面反射の分離, 研究報告自然言語処理, Vol. 2021-NL-248, No. 42, pp. 1–4 (2021).