






記事

多ピクチャーオブジェクト立体画像に対する適応的画像圧縮のベンチマーク評価

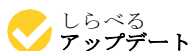
Alessandro Ortis ^{1,*} , Marco Grisanti ¹ , Francesco Rundo ² and  Sebastiano Battiato ¹ 

¹Department of Mathematics and Computer Science, University of Catania, 95125 Catania, Italy; marco.grisanti@phd.unict.it (M.G.); battiato@dmf.unict.it (S.B.)

²STMicroelectronics, ADG Group-Central R&D, 95121 Catania, Italy; francesco.rundo@st.com

*対応 : ortis@dmf.unict.it

4.0/) の条件に基づいて配布されているオープンアクセス記事です。



引用元Ortis, A.; Grisanti, M.; Rundo, F.; Battiato, S. A Benchmark Evaluation of Adaptive Image Compression for Multi Picture Object Stereoscopic Images (マルチピクチャーオブジェクト立体画像の適応的画像圧縮のベンチマーク評価) *J. Imaging* **2021**, *7*, 160.
<https://doi.org/10.3390/jimaging7080160>

アカデミック・エディターロマン
・スタロソルスキー、クオ・リヤ
ン・チョン

受領しました。2021年7月5日
受理されました。2021年8月18日
発行：2021年8月23日

出版社からのコメント：MDPIは、出版された地図や機関提携における管轄権の主張に関して中立的な立場をとっています。



著作権：© 2021 著者によるものです。
ライセンス：MDPI, Basel, Switzerland. この記事は、クリエイティブ・コモンズ表示 (CC BY) ライセンス (<https://creativecommons.org/licenses/by/>)

概要：立体視画像は、同じ被写体を異なる2つの視点から撮影した2枚の画像から構成される。この2枚の画像には多くの冗長情報が含まれているため、画質を維持したまま保存容量を削減することを目的とした新しい圧縮手法やデータフォーマットが次々と提案されている。マルチピクチャー画像符号化の標準は、MPO フォーマット (Multi-Picture Object) に代表されます。従来の立体視画像圧縮手法は、2つのビュー間の視差マップを計算し、残像とともに2つのビューのうちの1つを保存するものであった。適応型立体視画像圧縮と名付けられた別のアプローチでは、2つのビューだけを異なる品質係数で独立して符号化します。そして、2つのビューの間の冗長性を利用して、低品質の画像を向上させる。本論文では、立体画像圧縮の

問題を、圧縮データの標準的なフォーマットを得ることができる適応的立体画像圧縮のアプローチに焦点を当てて提示する。本論文では、解像度や撮影手法の異なる60のステレオペアを含む大規模かつ標準的なデータセットに対するベンチマーク評価を行う。本手法は、圧縮量、マッチング、最適化手法を変化させることにより、16種類の設定値で評価される。また、適応的アプローチは、他のMPOに準拠した方法と比較される。また、116人を対象としたHVS (Human Visual System) 評価実験により、デコード画像の知覚品質を検証している。

キーワード：立体視、立体視画像圧縮、マルチピクチャーオブジェクト、画像符号化

1. はじめに

立体視画像 (ステレオペア) は、2台のカメラで同じシーンを異なる視点から同時に撮影した左右一対の画像で構成される。この撮影は、人間の視覚システム (HVS) の両眼視をエミュレートすることを目的としています。このため、2台のカメラ間の距離は人間の目の瞳孔の間隔に設定されています。その際、左目には左の映像だけを、右目には右の映像だけを見せるように、適切なフィルターやデバイスを使用します。こうして、HVSに提示された2つの単一視点は、脳の視覚野で統合されるため、視聴者は本物の3Dシーンを知覚することができるのです。

立体視画像のストレージは、単一画像に必要なストレージの少なくとも2倍であり、それゆえ、立体視圧縮技術のアプリケーションの研究の動機となっています。MPO (Multi Picture Object) は、CIPA (Consumer & Imaging Products Association) が定義するマルチビュー画像のエンコードに用いられる標準化されたファイルフォーマットです[1]。このフォーマットは、JPEG エンコードされた画像の連鎖を一意のファイル (すなわち MPO ファイル) に適切に具現化し、マルチビュー取得のコンテキストで単一画像を復元し、それらを相互に関連付けるのに役立つ一連の追加情報を実装しています。立体視画像の場合、MPO ファイルには2つの画像のみがエンコードされます。しかし、このフォーマットでは、無制限の数のビューをエンコードすることができます。さらに、マルチビューを再構築するために必要なデータ

の画像は、最初の画像の JPEG メタデータフィールドを利用して、画像のメタデータに格納されます。そのため、MPOフォーマットでは、フィールドの追加やデータフォーマットの変更は必要ありません。

JPEGパイプライン[2]とその拡張[3]は、品質とメモリ要素の両方を考慮した画像パレットの最適な符号化を可能にすることで、デジタル画像圧縮のための標準を表しています。JPEGエンコードプロセスに基づき、パイプラインの1つまたは複数の要素を最適化する試みがいくつかなされています。例えば、[4]は、適切な量子化テーブルを考案するための代替戦略を提案しました。この方法は、既存のJPEG圧縮方式の圧縮性能を向上させることができることを実験的に示している[2,5]。

MPO は多視点画像の規格を提供し、いくつかの実用的な利点をもたらす。しかし、ステレオ画像圧縮の問題に取り組む多くの最先端技術は、圧縮画像の標準化を考慮しない符号化技術を実装しています。

既存のアプローチのほとんどは、2つのビューのうちの1つ、視差マップと、もう1つのビューの残差版を保存する[6-8]。そして、残差画像と視差マップを用いて、ある程度の信頼性をもって、圧縮されたビューを復元する。この手法では、視差マップや残像などの追加データを保存する方法が必要である。多くの場合、視差マップにはエントロピーベースの符号化が、残差画像にはDCT (Discrete Cosine Transform) が用いられている。特に、[8]では、左右の画像から計算された局所的な1次元高速フーリエ変換 (FFT) から立体視視差マップを推定することが提案されている。視差マップに基づくアプローチでは、再構成の際に問題が生じる。特に、視差マップでは通常、緩やかな遷移が観察されるが、オブジェクトのエッジによってパラメータが急激に変化することがある。この場合、視差のない画素（すなわち視差の低下）と、二重割り当てによるマッチングの問題が発生する[8]。このため、[8]では、左画像、右画像、視差マップに加え、エラー画像を定義し、ステレオペアのエンコーディングパイプラインに追加しています。

Schenkelら[9]は、ステレオペアを構成する2つのビューのジョイントデコーディングアプローチを提案しました。この提案手法は、JPEGパイプラインで圧縮された画像ペアのエンハンスメントを行うものです。しかし、画像の一部の領域は再構成することができません。また、JPEG 圧縮の品質が中程度の場合、PSNR (Peak Signal to Noise Ratio) が低下し、ゴーストが発生することが実験により確認されています。

[10]では、ブロックサイズと圧縮画像から計算される視差マップの品質を共同で最適化する、立体画像用の可変サイズブロック符号化アルゴリズムを提案しています。特に、視差マップの符号化に使用される、より詳細な画像領域に対して、きめ細かいピクセルブロックを適用しています。この手法は、ステレオペアのビットレートを下げながら、高い再構成品質を得ることを目的として設計されています。[10]で得られた実験結果は有望ですが、エンコーディングのためのブロックレイアウト定義に加え、ブロックレイアウトを記述する木構造、ブロック長マップ、ステレオペアのデコーディングに必要なその他の情報からなる追加データによる計算努力のオーバーヘッドが必要です。これらのデータ構造は、画像ペイロードの他に、アドホックなファイル形式で適切に保存される必要があります。

Poolakkachalilらは[11]で、カーブレット変換を用いた算術符号化方式による対称的な立体画像圧縮のアプローチを発表し、Stereoscopic Image Compression using Curvelet and Arithmetic Coding (SICCAC) と名づけている。さらに[13]では、2つのステレオ画像の差分と平均を量子化し、適応的算術符号化 (S2ICAC) を用いて符号化するパイプラインが提案されています。また、エントロピーに基づく符号化を用いて立体画像を符号化する方法もあり、[14]では、Stereoscopic Image Compression using Huffman Coding (SICHC) として評価されています。

上記の著作の多くは、平行カメラ光軸ステレオ方式による立体視画像の取得を前提としており、輻輳軸の設定は無視される。カメラ光軸の収束は、垂直方向の視差を発生させるパースペクティブ効果を伴う。平行カメラ光軸では、この垂直方向の視差が除去される。あるビューを他のビューから完全に再構成しようとするアプローチは、[8]で議論されているように、完全な再構成が不可能であることを考慮していない。また、実際のシーンでは、それぞれの画像上に、もう一方の画像では見えない領域が存在します。このように、立体視画像の圧縮問題を実画像・高画質画像という文脈で考えた場合、シーンの一部を無視したアプローチは考えられず、保存（ファイル）形式や伝送形式の標準化に関する問題にも取り組まなければならないのです。

[15]の研究では、立体視画像圧縮のためのエンドツーエンドのディープニューラルネットワークが紹介されています。具体的には、左右の画像間のホモグラフィ H を推定し、 H を利用して左の画像を右のビューにマッピングし、残像を保存する。提案アーキテクチャは、ホモグラフィ行列を推定するための深層回帰モデル、2つのオートエンコーダ、およびCQE（Cross Quality Enhancement）ネットワークを含んでいます。[15]の手法は興味深い結果を得ることができるが、データの質と量、計算時間、特定のハードウェアの観点から、非常に高いリソースを必要とする。また、入力画像に特定の解像度を強制的に与えるだけでなく、入力画像に偏りを持たせている。

は、近景（InStereo2K）[16]と遠景（KITTI）[17]のステレオペアに分けられた学習データセットに関連するものである。特に、[15]の手法は、 H 行列の推定にディープニューラルネットワークを利用しており、従来のジオメトリベースの手法では、左右のビューの対応関係が数個あれば、簡単に推定することができました。エンコード中、左の画像は H によって幾何学的に変換され、その後、2つの画像は2つの別々のオートエンコーダによってさらに処理される。このような処理により、エンコード時に推測される幾何学的制約が崩れる。このため、CQEネットワークが必要である。著者らは2つのディープモデルを提案し、前者は50.5M、後者は69.3Mのパラメータを必要とする。このモデルは、それぞれ約2000例を含む高品質立体画像の2つのデータセット [16,17] で学習されている。

[18]の著者らは、立体視画像の冗長性を利用した立体視ペアのビットレート低減に関する予備的な実験を行った。特に、この論文では、2つのアルゴリズムを組み合わせています。1つは低・中域のビットレートで、もう1つは中・高域のビットレートでより良い性能を発揮します。一つは、ラグランジュ乗数を用いて、ビットレートと歪みのトレードオフを改善するために視差マップを反復的に修正するものです。もう1つは、JPEGが補償精度をどのように扱うかについての単純化されたモデルに従って、ブロック単位で各視差を選択する。しかし、[18]の主な貢献は理論的なものであり、実際、提案されたアプローチは2つの立体画像でテストされたただけであった。

[19]において、著者らは、非常に低い品質損失でステレオペアの保存に必要な容量を大幅に削減する、MPO 画像圧縮の戦略を提示しました。この方法の主な利点の1つは、圧縮フェーズで、MPO準拠の圧縮ファイルを得ることができることです（図1参照）。復号化段階では、適切な復元段階を経て、2つの画像のMPO規格の復号化後に元の情報を復元します。[20]の論文は、[19]の研究を改良し、同じ再構成品質を維持しながら、全体の再構成段階の複雑さを劇的に減少させたものです。従来の方法と比較して、[19,20]の方法は、MPOフォーマット用に明示的に設計されており、撮影デバイスに直接実装できる適切なコーディング/デコーディングパイプラインを定式化しています。そのため、標準化をサポートし、パラレルまたはコンバージェントのステレオカメラシステムを使用して取得された立体視画像に関係なく動作します。しかし、この実験は、選択された少数のステレオペア（すなわち、23組）に対して行われたものである。しかし、[20]の手法は、[19]の手法を計算コストの面で大幅に改善することができた。

本論文では、実シナリオにおける立体映像のアルゴリズム評価のために設計された、標準的でよく知られたデータセットから取得した大規模な画像セットに対して、[20]の拡張ベンチマーク評価を実施する。さらに、再構成された画像の品質に関する主観的評価の結果を示す。この評価では、116人の参加者がそれぞれ10枚の画像を評価し、考慮されたデータセットからランダムに選択した画像について1160枚のHSVベースのテストのセットを作成した。そして、[20]の手法を[11,13]で使用したデータセットで評価も行った。

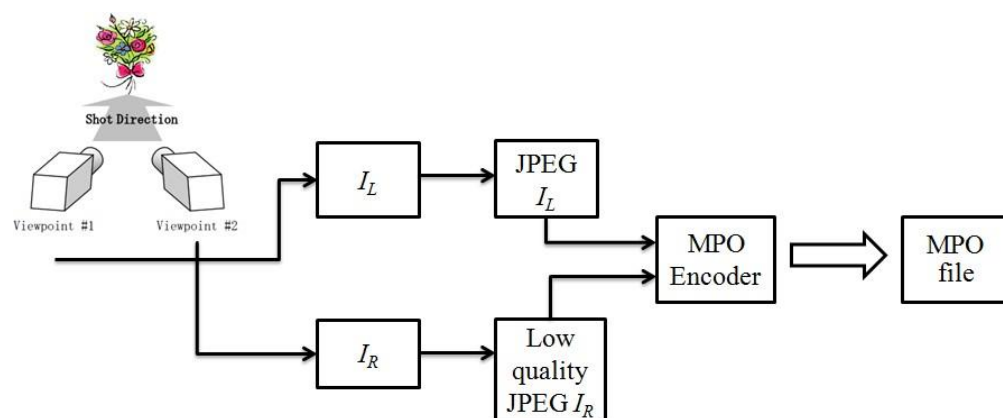


図1. エンコードパイプライン。各ビューは、異なる品質係数を用いて、MPOフォーマットに従って（すなわち、JPEG圧縮を適用することによって）符号化される。

本論文の残りの部分は以下のように構成されている。セクション 2 では、最新の手法の長所と短所を比較することで、今回紹介するベンチマークの動機付けを明確に述べています。セクション3では、採用した符号化／復号化パイプラインを詳述し、[19,20]で提案されたアプローチとの相違を比較する。また、2つのアプローチを詳細に説明し、計算量の観点から比較する。セクション4では、異なる解像度や撮影設定（平行カメラや収束カメラ）を持つ公共の標準化データセットから取得した60のステレオペアで16種類の実験設定を考慮した適応型立体視画像圧縮アプローチの実験評価を示す。また、他のMPO準拠の手法との比較評価や、再構成画像の知覚品質を評価することを目的としたHSVベースの実験も示しています。第5章では、本論文の結論を述べる。

2. 動機

多視点映像においては、MPOフォーマットが長年にわたって確立された標準であり、現在、ゲーム機を含む複数の機器メーカーが、複数のレンズシステムを備えた写真カメラで撮影した立体映像や3D映像を保存するために採用しています。このような背景から、[11,13,19,20]で提案された手法は、MPOフォーマットとの互換性を維持しつつ、撮影設定（平行軸/収束軸、画像解像度など）に依存しない符号化/復号化パイプラインを定義しています。前節で述べたように、他の方法では、ステレオペアを格納するための追加ペイロードやアドホックなデータ形式が必要であったり[6-9]、ゴーストのアーティファクトが発生したりすることが多い[9]。画像エンコード/デコードパイプラインの標準化は、特にそのような手法が撮影/レンダリングデバイスに組み込まれ、限られたリソースで高い性能を必要とする場合に、いくつかの利点をもたらします。数百万のパラメータと複雑な非線形演算を必要とする非常に深いニューラルネットワーク（DNN）に基づく方法は、そのようなデバイスに組み込むことはできません[15]。さらに、DNNは数百万のパラメータを持ち、それぞれが複雑な相互関係を持ちます。このように、ディープラーニングモデルはブラックボックスであると批判されてきたが、決定論的で説明可能な幾何学に基づくアプローチとは対照的に、完全に透明で、達成したソリューションがトレーニング環境の外で機能するかどうかを直接観察することができる[21]。実際、従来のComputer Vision技術では

niquesは、学習時間、処理、データ要件の削減から、特に幾何学関連分野[21]（例：動きからの構造、パノラミックスティッチングなど）に適用されるDNNよりも多くのアプリケーションで好まれます。ベンチマーク評価で評価された手法は、比較評価で検討された手法と同様に、MPO画像用に特別に設計されたものである。比較した手法はすべてMPOに準拠しており、低リソース機器に既に実装されている符号化・復号化パイプラインを再設計するため、容易に機器に組み込むことができる。また、これらの手法は、ステレオ撮影の設定（カメラ軸が平行か収束か）にも依存しない。しかし、これまで発表された実験は、画像の量、質、ばらつきに限界があった。さらに、他の方法との比較や、知覚された画質の主観的な評価も提示されていない。以上のことから、標準的で大規模なベンチマーク評価の必要性を感じた。

本論文で紹介するベンチマーク評価には

- 2001年から2014年のバージョンを含むMiddlebury-scenesデータセットの60個のステレオペアを用いた実験。
- 解像度は375 450から2016 2960の範囲で変化します。
- 異なる特徴検出と形状推定を組み合わせることで、合計16種類の評価設定を実現。
- 最近発表された他の5つの手法と、同じ公開データセットで、ビットレート節約と再構成品質の両方を考慮した比較評価。
- テスト回数が多く、参加者の母集団のばらつきが大きい場合に行う主観評価では、再構成された画像が高品質なものと区別がつかないことを確認しました。

3. 評価対象パイプライン

以下の段落では、符号化/復号化戦略について述べる。特に、[19,20]で実装された2つの異なるマッチングアプローチについて詳細に説明し、計算量の観点から比較します。

3.1. エンコードパイプライン。非対称圧縮

符号化段階では、提案手法は1つの画像ビューを低いJPEG品質係数で符号化します[2]。エンコード時には、2つの画像間の冗長性を利用し、高画質な画像を参照として低画質な画像を強調します。強調処理では、2つのビューから適切に抽出された画像ブロックを比較します。冗長性が検出されない領域（低画質と高画質の差が大きい領域）では、低画質のパッチからの情報のみが考慮されます。そのため、結果として得られる強調表示には一定の損失が生じます。しかし、実験によると、この損失率は数値的に無視できる程度であり、視覚的に知覚できるものではありません。JPEGのエンコードパイプラインは、圧縮量を制御するために品質レベルを定義しています。低品質の画像はより小さなJPEGファイルを生成し、高品質の画像は比較的大きなファイルを生成します。品質レベルは、JPEGエンコードパイプラインで使用される量子化テーブルを決定し、これらのテーブルは圧縮時の損失量を制御し、したがって、生成されたファイルのサイズを制御します。したがって、品質レベルは画像の視覚的品質とファイルサイズに直接影響します。

JPEGの圧縮量は、通常、品質レベルに対するパーセンテージで測定されます。一般的に、品質レベルが90%以上の画像は高品質、80-90%は中品質、70-80%は低品質とみなされます。70%以下の品質値で圧縮された画像は、一般的に非常に低品質であると言えます。このような品質レベルでは、エッジがシャープでなくなり、圧縮アーチファクトが見えるようになります。このような理由から、実験では、70%と65%の品質レベルを考慮して低品質画像を圧縮しました。

図1は、2つの画像ビューを異なる方法で符号化する、提案する符号化パイプラインの詳細を示しています。低画質ビュー I_R の各画像ブロックは、高画質ビュー I_L を利用して再構築されます。

私はPEGの高画質レートで L 、右の画像は低画質レートで R 。しかし、右と左のビューの役割を逆転させれば、同じパイプラインを適用することができます。

3.2. NCCベースの復号化アプローチ

[19]で提案されたアプローチは、画像パッチ間の相関に基づく画像ブロックマッチング法を実装しています。画像 I 内の与えられたパターン t を照合する一般的な方法は、 x 方向と y 方向にそれぞれ u と v ステップだけシフトしたテンプレート t の可能な各位置 (u, v) で計算される正規化相互相関(NCC)スコア $ncc(u, v)$ を考慮することである。NCC係数は以下のように定義される。

$$ncc(u, v) = \frac{\sum_{x,y} \{ [I(x, y) - I_{u,v}] [t(x - u, y - v) - t] \}}{\sqrt{\sum_{x,y} [I(x, y) - I_{u,v}]^2 \sum_{x,y} [t(x - u, y - v) - t]^2}} \quad (1)$$

ここで、 $I_{u,v}$ はテンプレート t の領域内にある画素 $I(x, y)$ の平均値を (u, v) でシフトしたもので、 t は t の画素の平均値である。

デコード時には、左右のビューが重ならないように細分化されます。各ブロックに対し、左画像と右画像から抽出したブロックの冗長性を利用し、低品質のブロックから抽出したブロックの品質を向上させることを目的とした復号化手順がある。復号化アルゴリズムは、各チャンネルから抽出された各ブロックに適用される。

マッチングアプローチ

低品質画像 I_R から抽出された一般的な i を bR と呼ぶと、その目的は、 bR を強化するために冗長性を利用できる高品質画像 I_L の最良の部分画像を見つけることである。この目的のために、このアルゴリズムでは2つのブロック候補を考慮します。

- bR と同じ位置にある I_L の $N \times M$ ブロックである。
- 正規化相互相関を計算して得られた $N \times M$ ブロックの I_L (NCC) [22]を、 I_L のすべての $N \times M$ 個の部分画像 bR の間に適用し、NCC値が最も高い部分画像を考慮します。

実際には、 I_L 、 bR のすべての可能な位置についてNCCを計算する必要はない。最適化されたアプローチは、 bR の元の位置を考慮して I_L の限定された領域でNCCを計算することである。

そして、 bR との絶対値差の和(SAD)を最小にする候補ブロックを選択する手順となる。選択されたブロック i は、拡張手順によって利用されるは、3.4節で説明した。

3.3. 幾何学に基づく復号化アプローチ

[20]では、[19]の研究を拡張し、マッチング段階の有効性の向上と計算コストの最適化に焦点を当てたアプローチを示している。この2つの目的は、ステレオスコーピックペアの幾何学的制約を利用することで達成された。

3.3.1. エピポラ幾何学

Epipolar geometryは、立体視画像において、同じ3Dシーンを表現する2つの画像間の特性や幾何学的な関係を記述する。このような幾何学は、2つのカメラのパラメータとその相対位置によって完全に記述されるため、シーン構造から独立しています[23]。

定義1 (基本行列). F と表記される基本行列は、対応する任意の点 x_1 と x_2 の組に対して、以下の条件を満たすランク2の 3×3 行列である。

$$x_1^T F x_2 = 0 \quad (2)$$

ここで、 x_1 は一方の画像ビューの点であり、他方の画像ビューの x_2 に対応する。

基本行列は、シーンに対する撮影設定（すなわち、2つのカメラの相対的な位置と向き）の本質的な幾何学的特性を符号化します。実際、式 (2) の x_1 と x_2 は、2つのカメラ画像平面上の同じ実3D点の投影に対応します。したがって、立体画像の左右の視点から撮影された2点が同じ実世界の点を表現しているとすれば、それらは式 (2) の関係を満たす必要があります。この性質を利用し、[20]で提案された手法では、左右のビューの対応関係の探索範囲を大幅に縮小している。したがって、マッチング段階での計算量を削減することができます。実際、点線二元論定理によれば、ある立体画像が与えられたとき、第一ビューの任意の点 x_1 に対して、第二ビュー上に対応するエピポーラ線 l_2 が存在する。さらに、 x_1 に一致する第2ビューの任意の点 x_2 は、 l_2 上にある。実際、 l_2 は、点 x_1 から第2ビューへの光線の投影であり、第1カメラの中心を通過している。

を利用することで、 x_1 を第二画像上の対応するエピポーラ線 l_2 に写像することができる。は以下の式で表されます。

$$l_2 = Fx_1 \quad (3)$$

つまり、基本行列 F によって、1枚目の画像にある任意の点 x_1 と、それに対応する点 x_2 を含む2枚目の画像のエピポーラ線 l_2 との間に直接的な関係を持つことができる。したがって、点 x_2 の探索を l_2 の点に限定することができるのである。

3.3.2. 画像のブロック化

3.2 節で説明した復号化手法では、低画質ビューは重なり合わない多数のブロックに細分化される。ブロックサイズは入力画像の大きさに応じて適応的に選択され、画像領域をすべてカバーすることを目的とし、結果として少数のかなり大きなブロックになる。このようなブロックサイズでは、各画像は約12ブロックに細分化されます。このようにブロック数を少なくすることで、復号化アプローチを合理的な時間で実行することができます。次に、低画質画像から抽出した各ブロックを、NCC計算に基づくテンプレートマッチング手法で高画質画像と比較し、最も類似した高画質領域を検出する。

このアプローチにはいくつかの限界があります。ブロックの寸法が大きいため、抽出された各画像パッチは、異なるオブジェクトや撮影シーンのいくつかの詳細を含む、シーンのかなり大きな領域（例えば、合計12ブロックの場合、各ブロックは画像の8%以上を含む）を描画します。2台のカメラの視点異なるため、各物体の見え方は左右で微妙に異なります。このような違いは、抽出されたブロックの大きさに応じて増大し、特に乱雑なシーンでは、より多くのオブジェクトが対応するパッチに含まれるようになる。さらに、この処理は各画像チャンネルに対して繰り返されます。その結果、照合処理が粗くなり、計算コストがかかる。

これに対し、[20]で示された方法は、非常に少数のブロックを考慮して画像を細分化する。すべてのブロックは同じ寸法であり、マッチング手順はブロックのすべてのカラーチャンネルに同時に適用されます。具体的には、半径 r を一定値（[20]では $r=20$ ）に設定し、低画質画像を $(r+1) \times (r+1)$ に分割する。

オーバーラップブロック必要であれば、画像の全領域をカバーするために、いくつかのオーバーラップするブロックは

を画像の右側と下側に定義した。ブロックの定義例を図2に示す。この例では、Middlebury 2003データセット[24]の画像「Cones」を110ブロックに細分化し、その中に重複するブロック（青で描かれている）を多数含んでいます。画像ブロックの次元を小さくすることで、処理されたデータのきめ細かな検索が可能になり、カメラシステムに対する複数のオブジェクトの位置が大きく異なるような、シーンの粗い部分の大きな行列を扱うことを回避することができるのです。例えば、同じ画像に対して、[19]で使用された方法は、サイズ360 360の12ブロックを定義するのに対し、[20]の方法は972ブロックを抽出し、より計算効率が高くなります。

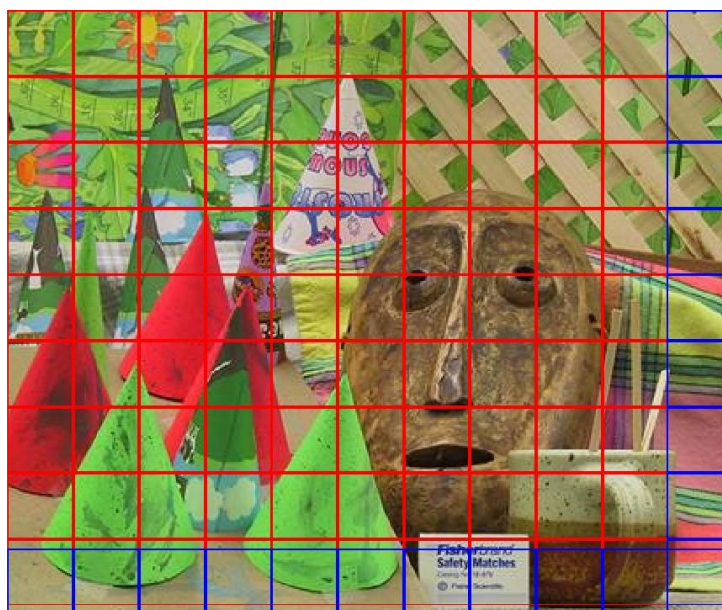


図2.20]で採用されたブロッキングスキーマの例。

3.3.3. マッチングアプローチ

ジオメトリベースアプローチの主な改良点は、マッチング戦略に関連している。以下では、従来通り、左側のビューを高画質画像とみなし、右側のビューを低画質画像とみなします。右側のビューから抽出されたブロック b_i があると、システムはブロックの中心 c_i を考慮し、式(3)で定義される点線関係を適用して、左側のビュー上の対応するエピポーラ線 l_i を計算する。エピポーラ幾何学(セクション3.3.1参照)によれば、左画像上の c_i に対応する未知の点は l_i 上にあることが分かっている。したがって、ブロック全体 b_i のマッチングは、エピポーラ線 l_i に沿ってブロック中心 c_i をシフトさせた左画像上の b_i の可能な位置に限定できる。したがって、画像ブロックのマッチングは、エピポーラ線上の c_i の可能な位置に限定できることが主な利益となる。このような位置の数は画像幅にほぼ等しいが、以下の段落で詳述するように、採用したアプローチは探索範囲をさらに狭める。抽出されたブロックの中心に関連するエピポーラ線は、式(3)を利用して計算されるため、基本行列 F の推定が必要である。

との対応関係から行列 F を推定することができる。は、左側と右側のビューで構成される。実際、2つの画像間のマッチング点の各ペアは、 F に対する線形制約を与える(すなわち、式(2))。その結果、基本行列 F は、8点アルゴリズムを適用して、少なくとも8つの独立したペアから線形に推定することができる[25]。8点アルゴリズムは単純で効果的であるが、入力に対応関係の精度に対して敏感である。代替案としては、Least Median of Squares推定があり、これは間違っただ対応関係の存在に頑健であるが、入力データの50%以上の数のマッチング点の良いペアが必要である。

8点アルゴリズムと最小二乗法推定はともに、多数の入力対応点を必要とする。これらの対応点の組は、SIFT (Scale-Invariant Feature Transform) [26] と Harris keypoints [27] という2つの異なる局所特徴検出のアプローチを考慮して計算されている。図3は、右画像(図3b)から取得した3つのサンプル点から、式(3)を適用して3つのエピポーラ線(図3a)を計算した例である。左図に描かれた点は、右図から取得したサンプル点の x と同じ x を持つエピポーラ線上の点である。実際、これらの点は右図の3つの点とは一致しないことがわかる。図4は、2番と3番の点の詳細を示している。

3: これらの点は右図（図4b）の円錐の頂点（点2）とマスクの鼻の先端（点3）上に置かれ、それらのエピポーラ線は左画像（図4a）の対応する点を正しく通過しています。また、これらの線上にある同じx座標の点（左図）は、対応する点に対してずれているように見える。このことから、右側画像の点が与えられた場合、左側画像上の対応する点がエピポーラ線上にあること（すなわち、式（3））以外に、この未知の点の位置は、右側画像で選択した点の同じxの点に近いことが示唆される。



図3 右画像IR右画像（b）のIR（1、2、3）の任意の点を考えると、[20]の方法は、そのエピポーラ線を計算することによって、左画像ILの対応する点を検索する（a）。



図4.図3に示した画像の詳細。円錐の先端（右図IRを描いた画像（b）の点番号2）は、左図ILの対応するエピポーラ線（（a）の線番号2）上に位置する点に対応する。

3.3.4. レンジリダクション

エピポーラ線を推定することで、探索範囲を一本の線にすることができます。つまり、画像ブロック b_i があれば、その中心 c_i のエピポーラ線を推定し、中心が i 上にある可能なパッチのみを考慮してブロックマッチングを行うことができる。それでも、次のことを観察すれば、比較の回数はさらに減らすことができる。

1. 探索した点のy座標は、エピポーラ線のy値で与えられる。
2. 検索された点のx座標は、右図の対応する点のxに近い。
3. 式(2)は、対応する任意の点の組で満たさなければならない。

検索範囲を最小にする目的で、[20]では以下のアプローチが考案されている。右側のビューから抽出されたブロックの中心点 x_1 が与えられると、式(3)を利用して、対応するエピポラ線 l_1 が計算される。次に、各点 $x_2 \in l_1$ 、 $x_1 \parallel Fx_2$ の値が計算される。式(2)より、 x_1 の対応点を左図で表すと、 $x_1 \parallel Fx_2$ の値はゼロクロス線を表すことになる。さらに、(1)によれば、 x_1 の対応点は、 $x_1 \parallel Fx_2$ の値がゼロとなる点の周辺に存在することになる。この領域にマッチング検索をさらに集中させるため、は、次のような関数が定義されています。

$$\varphi = (x_1 \parallel Fx_2)^2 \quad (4)$$

関数 φ は放物線を定義する。特に、 φ の頂点は0に近い。図5に φ の例を示す。 φ の値が頂点から離れるにつれて急激に大きくなる様子が観察できる。さらに、 c_i の x 座標が同じ点(図5では赤丸で表現)は、頂点の近くに配置されている。

実験では、2.3という値を設定することで、候補ブロックの数を大幅に減らすことができた(図5参照)。しかし、固定閾値のヒューリスティックを用いる代わりに、適応的な方法を適用している。これは、以下の適応的なルールに基づいている。

1. 閾値は2.3に設定されています。
2. c_i と同じ x の点が閾値処理後に得られた範囲に含まれない場合、これまでの考察により、この点が探索範囲に含まれるまで閾値を増大させる。
3. 検索範囲が空の場合、範囲が空にならなくなるまで閾値を0.5ずつ繰り返し増加させる。

これらのルールにより、効果的な適応閾値の定義が可能になった。

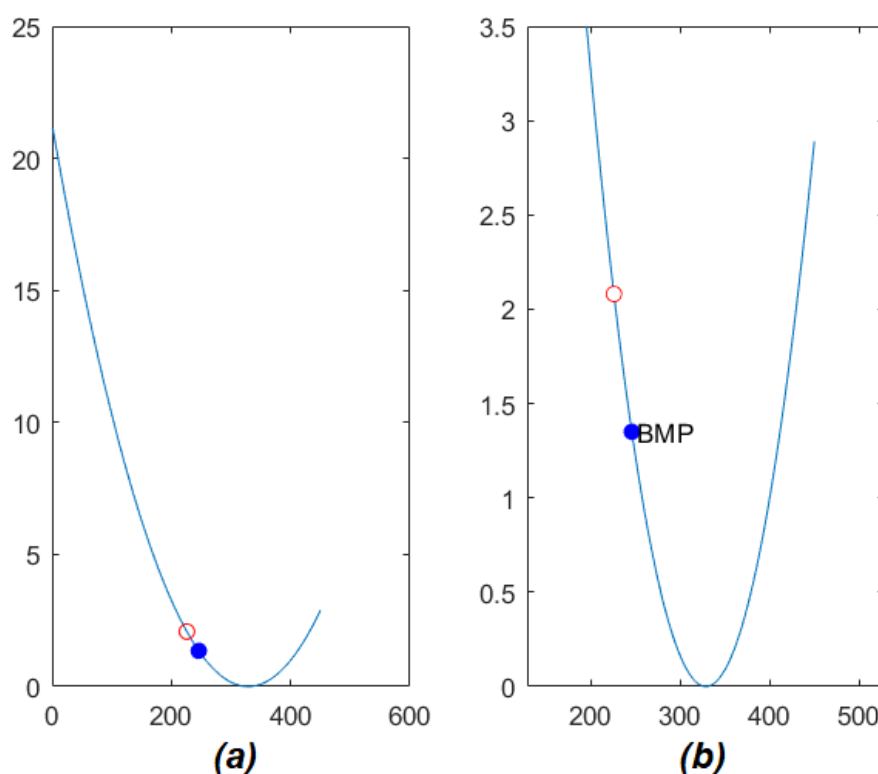


図 5.20]のアプローチを用いたマッチングの例 (a)、 φ 関数の頂点付近の詳細(式(4)参照)。赤丸は(b)の点と同じ x を持つ φ 上の点である。青い点は、アルゴリズムによるBMP (Best Matching Point) に対応する φ の値を表している。

3.3.5. ブロックマッチング

3.3.4節で述べたアプローチにより、探索範囲は大幅に縮小される。次に、右図のブロック b_i が与えられると、マッチング手順は、中心が l_i にあり、 x 座標が縮小された範囲にあるような候補ブロックを考慮して、 b_i の最良のマッチングブロックを検索する。これらの制約により、ブロック間の比較の回数を大幅に減らすことができる。ブロックの比較は、 b_i と選択された候補ブロックの間の二乗差の総和 (SSD) を計算することによって行われる。そして、最も低いSSDを達成したブロックが、強化フェーズで採用される (セクション3.4参照)。

実験によると、一様なブロックが存在する場合、上記の方法では偽陽性例を含んでしまい、マッチングが失敗する可能性があることがわかった。このような場合、閾値が高いと、幾何学的特性を考慮して計算された理論位置から離れたところにSSD的に最適なブロックが配置される可能性がある。つまり、画像に大きな一様領域 (壁など) がある場合、その領域から抽出された画像ブロックは、選択された画像パッチを特徴付けるエッジが少ないため、不一致ブロック間の差異をSSDが捉えることができず、互いに一致する可能性があるのです。

JPEG圧縮が低周波を保存する傾向があることを考慮し、[20]の方法は、低ピクセル分散のブロックをフィルタリングすることによって、拡張手順から均一ブロックを破棄する。この戦略により、比較の数 (すなわちSSDの計算) をさらに減らすことができ、文脈上、均一な候補ブロックの存在によって引き起こされるマッチングの問題を回避することができます。実験によると、この手法によって、結果は品質面で影響を受けず、時間性能は大幅に改善されました。したがって、このシンプルな選択により、品質を維持したまま、照合処理をさらに高速化することが可能となりました。

3.3.6. パーシャルマッチング

カメラの視点や向きの違いにより、一方の画像ビューの境界付近にある領域が、他方の画像ビューのブロックに対応し、部分的にしか表示されないということが起こり得ます。これは、シーンの一部が2つのビューのうちの1つでしか表現されていない場合に起こります。このような部分マッチングの問題に対処するため、[20]の手法では、より低次元の候補ブロックを許容しています。具体的には、処理されたブロックの寸法の60%以上の寸法のブロックをすべて含む手順である。

b_i と部分候補ブロックとの比較のため、 b_i の値を用いて欠損画素を埋める。

図6に、部分一致とブロック合成の例を示す。特に、1行目は立体視用画像「花1」のブロック番号111を部分一致させた例であり、2行目、3行目は立体視用画像「コーン」に関連する例である。図6の1行目は、ステレオペアの右側のビュー (すなわち、図6の1行目の (a) 列) にシャンデリアが見えており、3.3.2節のブロック化スキーマを考慮して、ブロック b_{111} に描かれていることを表しています。このオブジェクトは、左の画像ビューでは部分的にしか描かれていない。アルゴリズムが、両方の画像によって示されるシャンデリアの部分を表すブロック b_{111} に対してマッチング手順を実行するとき (図6a)、参照画像の左側に配置された41 25ブロックとの部分マッチングも考慮されます。図6bの青い部分は画像パッチ内の欠損部分を表し、これを同じ位置の b_{111} の画素で埋め、図6cに示すような合成ブロックを得ます。

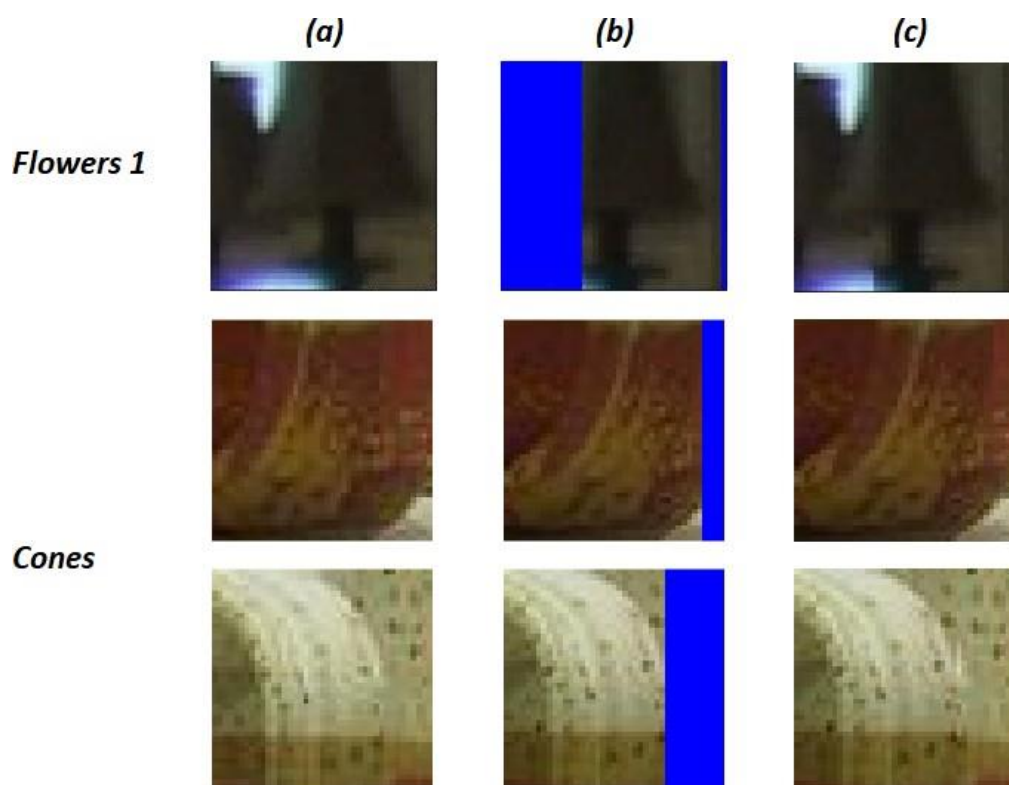


図6.部分一致ブロックでの克服のために行われたブロック合成の例。各行はブロック合成の例を示しており、1行目は画像「花1」のブロック番号111 (a)、部分マッチングブロック (b)、合成後のブロック (c) である。2行目と3行目は、画像「円錐」から抽出したブロックの部分一致の例で、それぞれブロック番号17と13に関連するものである。

3.4. イメージエンハンサー

エンハンスメントステップでは、Kohonen update rule [28]の簡略版に基づく以下の式を採用している。

$$b_{i \leq th}^R(u, v) = \begin{cases} \mathbf{f} \mathbf{b}^R(u, v) + \alpha \cdot d_i(u, v) & \text{if } d_i(u, v) \\ b_i^R(u, v) & \text{すなわち} \end{cases} \quad (5)$$

どこ

$$d_i(u, v) = b_i^R(u, v) - b_i^L(u, v) \quad (6)$$

ここで、 $b_i^R(u, v)$ は強調されたサンプル、 $b_i^L(u, v)$ はそのブロックが選択したサンプルである。

マッチング手順と $b_i^R(u, v)$ は、エンハンスメントが必要なサンプルです。このサンプルに

式(5)では、対応するサンプルのペア間の類似性に応じて、 b_i^R のいくつかのサンプルの値が b_i^L の対応する値に近づくように移動される。パラメータ α と th (我々の場合、 $\alpha=0$ 、25、 $th=0043$) は、式 (6) の $d_i(u, v)$ として定義される2つのサンプル値間の距離に基づいて再構成手順 [28] を制御する2つの係数である。式 (5) のパラメータは、3DMedia コレクションの 23 枚の MPO 立体視画像に対して以前に行われたグリッドサーチ評価 [19,20] によって、非常によく得られている [29]。本論文で紹介する拡張評価では、同じパラメータを使用しており、評価したアプローチの一般化能力をさらに評価している。

3.5. 計算複雑性

本節では、3.2節で述べたNCCベースのマッチング方法と3.3節で述べたジオメトリベースのマッチング方法について、計算機的な観点から比較を行った。

3.6. NCCベースの複雑性

サイズ $M \times N$ のシーン画像からサイズ $m \times n$ の参照テンプレートのマッチングを見つけるために用いられる正規化相互相関 (NCC) の計算複雑度は

$$O(mnMN) \quad (7)$$

したがって、NCCベースのマッチングのコストは、各 $m \times n$ ブロックについて、以下のようになります。

$$TNCC_{\text{matching}} = O(mnMN) \text{である。} \quad (8)$$

ブロック数を D とする。

$$D = \frac{MN}{\text{エムエヌ}} \quad (9)$$

したがって、3.2節で説明したマッチング手順による総コストは

$$\begin{aligned} TNCC_{\text{based}} &= D \times TNCC_{\text{matching}} = \text{となる。} \\ D \times O(mnMN) &= O(M^2 N^2) \text{となる。} \end{aligned} \quad (10)$$

3.7. 幾何学に基づく複雑性

サイズ $\times mn$ の2ブロック間のSSDの計算コストは $O(mn)$ であるから、3.3.3節で説明したマッチング方式によるコストは

$$T_{\text{GeometryMatching}} = k \times O(mn) \quad (11)$$

ここで、 k は最適なマッチングを見つけるために比較されるブロックの数である。したがって、 D をサイズ $m \times n$ の処理済みブロックの総数とすると、3.3節で説明した幾何学に基づくアプローチのマッチング手順の総コストは、以下のようになります。

$$\begin{aligned} T_{\text{GeometryBased}} &= D \times \\ T_{\text{GeometryMatching}} &= \text{。} \\ D \times k \times O(mnMN) &\text{である。} \end{aligned} \quad (12)$$

3.3.4節で説明した範囲縮小のアプローチにより、 k の値は非常に小さくなっている。特に、 $O(mn)$ という項に関しては、 k の値は無視できるほど小さい。したがって、総コストは

$$T_{\text{GeometryBased}} = D \times O(mnMN) = O(MN) \text{となる。} \quad (13)$$

エンハンスメントプロセスはNCCベースとジオメトリベースの両方で同じであり、その計算コストはマッチングコストより低いため、上記の詳細な計算分析では考慮されていない。実際、エンハンスメントプロセスのコストは線形であるため、次数の高いマッチングステップのコストに単純に加算することができる。したがって、比較されたマッチング手順の総コストのみを考慮し、その計算評価を行うことが可能である。以上の解析から、形状に基づく再構成は、二次関数的な成長（すなわち、 $O(M^2 N^2)$ ）から線形的な成長（すなわち、 $O(MN)$ ）に順位を下げて効率を向上させることができることがわかった。図7は、画像の解像度（すなわち、 $M \times N$ ）に対する計算時間（秒）の分布を示す。青い点は (MN, time) の組を表し、オレンジ色の点は $M \times N$ から一律ブロックチェックのために処理されなかった画素数を引いた後の同じデータを表しています。後者の点群（オレンジ色の点）は、 x 軸（すなわち、 $M \times N$ ）に対して直線的に分布していることが観察される。

実験は、Middlebury dataset 2014の画像群（つまり、最も解像度の高い画像群）を用いて行われた。

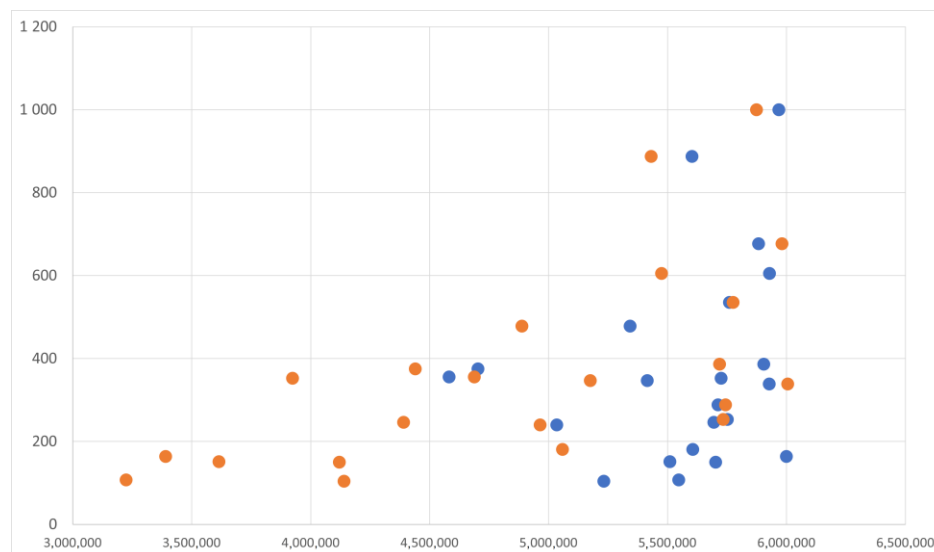


図 7.MxN (x軸) 対 計算時間 (y軸)。オレンジ色の点は、均一なブロックを無視した場合を考慮したものである。

4. 実験風景

提案手法は、[20]において、花の合成、自然風景、動物、建物など、実際の困難なケースから撮影された高解像度画像で評価することに成功しています。提案手法の評価結果は、ビットレートの節約とロッキーを考慮して、表1に示すとおりです。特に、データセットには、[1]規格に準拠した23枚の立体視用MPO画像と、異なる解像度（1440 1080、1620 1080、1444 1080、1924 1080、または1920 1080）。高画質画像のエンコードにはJPEG品質係数を85に設定し、低画質画像のエンコードには標準的な量子化テーブルを用いて65または70に設定しました。

この結果は、総ビットレートで圧縮利得を示し、品質損失はPSNR（Peak Signal to Noise Ratio）指標を考慮して測定される。特に、各MPO画像と、低画質画像の圧縮に使用したJPEG圧縮率（すなわち、65または70）の各値について、表1は、復号に使用したブロックの寸法（第2列）と、強調処理後に得られた再構成画像について計算した損失（dBの単位で）を報告しています。ビットレートの節約は、低品質画像のエンコードに必要なスペースと、85に等しい品質係数を使用して同じ画像をエンコードするのに必要なスペースを比較することによって計算されます（3列目と5列目）。なお、ブロックの大きさは、元画像の解像度に応じて、各ステレオペアにアドホックに選択されています。実験では、非常に低いロッキーで興味深いビットレート値を示しました。しかし、検討したデータセットには23個のステレオペアしか含まれていません。提案する実験では、構築され、一般に共有されている、よく知られた標準的なステレオスコーピックデータセットのプールを考慮しました。2001年以来、Middlebury Stereo Datasetsは、立体画像圧縮アプローチを含むいくつかの立体画像アルゴリズムを評価するために考慮されてきました。ここで紹介する実験では、2001年から2014年の間に公開されたMiddlebury Stereo Datasetのすべてのバージョンから60の立体視ペアを検討しました。実際、画像の解像度と採用された撮影技術によって異なるいくつかのバージョンがあります。特に、我々は検討した。

- Middlebury-scenes 2001 [30]の区分的平面シーンの5つのデータセット。
- Middlebury-scenes 2003の2つのデータセット[24]。
- Middlebury-scenes 2005の9つのデータセット[31]。
- Middlebury-scenes 2006の21個のデータセット[32]。
- Middlebury-scenes 2014の高解像度シーンの23のデータセット[33]。

表1.19,20]で報告された最初の実験で得られた結果。

MPO画像	N × M	低品質		65低品質 70	
		ロッシー	ビットレート保	ロッシー	ビットレート保
Flowers1	360 × 360	(dB)	存	(dB)	存
Flowers2	360 × 481	2.17	40.70%	1.65	34.60%
花3 キ ヤ ッ ス ル ド ー ム ペ リ オン	360 × 481	2	40.60%	1.32	34.50%
		2.76	48.70%	2.76	41.70%
		2.62	38.30%	2.18	32.50%
		2.73	37.10%	2.64	31.10%
ホールウ エイスタ チュー 図書館 ホール ガーデ ニング ブリッ ジ	360 × 483	2.34	37.60%	2.34	31.80%
		360 × 360			
		2.33	37.60%	2.33	32.10%
		2.59	41.90%	2.6	35.70%
Autumn1 Autumn2 Autumn3 Autumn4	360 × 361	1.98	38.70%	1.71	32.70%
		1.66	41.10%	1.4	34.90%
		2.11	39.50%	1.82	33.50%
		2.73	35.30%	2.53	29.80%
1 どうぶ つ	360 × 361	2.6	36.40%	2.4	30.60%
		2.41	36.50%	2.15	32.50%
		2.38	37.00%	2.15	31.20%
		2.62	36.80%	2.44	31.20%
どうぶ つ	360 × 361	2.85	36.10%	2.44	30.40%
		2.69	37.40%	2.52	31.40%
		2.67	36.90%	2.33	31.10%
		2.16	38.80%	2.16	32.80%

提案する実験に使用される3つのステレオペアのベンチマークセットでパイプラインの大規模な評価を行うことができ、また、異なる解像度の入力に対するパイプラインの性能を比較することができます。このアプローチは、画像点の対応関係の初期セットを定義するためのSIFTまたはHarrysキーポイント検出と、基本行列推定のための8点アルゴリズムまたは最小二乗最適化を考慮し、局所特徴検出と基本行列推定アルゴリズムのいずれかを变化させて評価されています。予想通り、SIFTアルゴリズムは、いくつかの外れ値が存在するものの、大量の相関を提供します。一方、Harrys keypoints法は、対応点数は少ないが、外れ値の存在率は低い。セクション 3.3.3 で説明したように、基本行列を推定するための最適な手法の選択は、2つの画像ビュー間の利用可能な点の対応付けの数と質に依存します。

本論文では、紙面の都合上、Middlebury- scenes 2014 [33] と Middlebury-scenes 2006 [32] のデータセットで得られた結果のみを掲載する。本論文では、紙面の都合上、Middlebury-scenes 2014 [33]とMiddlebury-scenes 2006 [32]の2つのデータセットで得られた結果のみを掲載する。特に、Middlebury-scenes 2014データセットでの実験結果は表2（低品質係数65）と表3（低品質係数70）に示されており、一方、Middlebury-scenes 2006データセットで得られた結果は表4と表5に報告されています。これらの結果から、適応的手法により、少ない計算量で高い画質を得ることができることが確認されました。いくつかの実験

メンツは、10-4dB以下の再構成損失を得ることができました。実験結果を観察することで

の結果、他を強く凌駕する設定は存在しない。しかし、ロッシーと計算時間の平均的なパフォーマンス（各表の最後の行に報告）を考慮すると、基本行列を推測するための最小二乗平均最適化に基づく方法は、ロッシーの点で優れたパフォーマンスを示し、8ポイントアルゴリズムに基づく方法は、わずかに時間のパフォーマンスが優れていることが示された。

.384362.172012.171892.592572.59 887

Average401.553751.284531.584611.62596

聖徳太子

2.020.812960.861790.913670.82 446

表3.middlebury-scenes 2014データセットで、低品質70を考慮した結果。

[illegible]

表4. Middlebury-scenes 2006データセットで、65の低品質を考慮した結果。

ステレオペアサイズブロック数			方法											
			低品質65											
			ビットレ 節約率(%)	LmedS				Norm8Points						
				ハリス		シフト部		ハリス		シフト部				
ID			ロッシー (dB)	時間 (秒)	ロッシー (dB)	時間 (秒)	ロッシー (dB)	時間 (秒)	ロッシー (dB)	時間 (秒)				
1	アロエ555 × 644	6203	224	37	1.58	6	2.01	11	1.83	4	2.01			
2	Baby2555 × 620	555 × 6565	224	40	0.87	4	1.24	6	1.65	26	1.24			
ベビー3			4238401.61361235131.0261.3913		1.00	10	1.59	5	1.91	27	1.59			
ボウリング1		555 × 6266		238	39	1.81	9	2.27	15	1.81	13			
布1	ボウリング2555 × 6268	555 × 665	7224361.52111277161.77111.386		1.75	22	1.65	9	1.85	20	1.65			
布2		555 × 6												
布3		555 × 62610		224	38	1.85	9	1.85	15	2.18	24			
布4		555 × 65011		224	37	2.18	27	1.84	8	1.91	10			
	植木鉢555 × 656			238	39	1.59	16	2.01	12	1.59	7			
	ランプシェード1555													
		×650	12224393.18143.18131.9193.1813											
	ランプシェード2555													
		×650	13224392.0182.0162.01142.016											
	Midd1555 × 698													
			14252361.4952.42311.4982.4231											
	Midd2555 × 683													
			15238361.2751.27101.2741.2710											
モノポリー	555 × 665													
			16238361.2781.94261.9451.9426											
	プラスチック555													
		×635	17224340.4290.42130.4270.4213											
	ロックス1555 × 638													
			18224381.6582.40131.65152.4013											
	ロックス2555 × 638													
			19224381.6592.0582.0552.058											
	ウッド1555 × 686													
			20238420.59150.5961.33290.596											

J.イメージング2021, 7,160					2024の
21Wood2555	×	653224401		.42310.428<10-460	.42 8
				Average381.53131.68121.57131.68	
		12			
	聖徳太子				
		1.680.598.730.696.500.548.110.69	6.50		

表 5.Middlebury-scenes 2006データセットで、70の低品質を考慮した結果。

ID	ステレオペアサイズブロック数			ビットレ 節約率(%)	方法							
					低品質70							
					LmedS				Norm8Points			
					ハリス		シフト部		ハリス		シフト部	
					ロッシー (dB)	時間 (s)	ロッシー (dB)	時間 (s)	ロッシー(dB)	時間 (s)	ロッシー (dB)	時間 (s)
1	555 × 620	224 × 224	2	32	1.26	9	1.45	18	1.52	8	1.52	8
2	555 × 620	224 × 224	2	34	0.56	1461	1.82	341	1.32	2861	0.87	1.02
3	555 × 656	224 × 224	2	32	0.56	9	1.45	18	1.52	8	1.52	8
4	Baby	3238340.9161.17400.91160.91		11								
5	Bowling	1224331.00101.00261.91161.00		5								
6	Bowling	2238332.27301.40141.4072.08		8								
7	Cloth	1224311.83281.47291.4791.71		13								
8	Cloth	2224321.32131.21151.75101.65		7								
9	Cloth	3224321.50101.50161.6961.50		10								
10	Cloth	4224312.05251.49131.49101.78		8								
11	Flowerpots	238341.5991.59222.01162.42		19								
12	Lampshade	1224332.33112.33142.59262.33		8								
13	Lampshade	2224342.0182.0182.0182.33		30								
14	Midd	1252311.27111.27111.2781.49		9								
15	Midd	2238310.6840.68140.6841.49		29								
16	Monopoly	238301.1051.94291.27151.94		20								
17	Plastic	224291.42220.4260.42180.42		8								

J.イメージング2021,										2224の
7,160										
18Rocks1224322.27331.3392.40371.65	555 × 638	7								
19Rocks2224321.5090.9161.6591.65	555 × 638	6								
20Wood1555 × 686238350			.5616<10-4121			.3329<10-48				
21Wood2555 × 653224331					.42250.4210<10-461	.42	22			
										Average321.47151.26181.46141.48
										12
標準偏差										
										1.450.538.980.559.890.619.090.62 7.40

4.1. 比較評価

[11,13,14]のアプローチは、量子化後に適用する符号化・復号化変換、すなわち、それぞれ、カーブレット、算術符号化、ハフマン符号化に大きな違いがあるが、MPO標準[1]と同様のパイプラインである。このような符号化・復号化パイプラインは、ステレオペアの取得設定に依存しない。これらのアプローチは、テキサス大学の有名なLIVE 3D画質データベース[34]で評価されている（LIVE 3D画質データベースは、http://live.ece.utexas.edu/research/quality/live_3dimage_phase1.html（2021年6月21日にアクセス））。このパラグラフでは、同じ評価設定を考慮して、適応的アプローチ[20]の[11、13、14]に対する比較評価について報告する。表6は、本ベンチマークに含まれるすべての実験設定を考慮して、LIVE 3D データセットに適用した適応手法の詳細な結果である。実際、すべての実験設定は、左画像に同じエンコード/デコード・パイプラインを適用し、右側のビューには特定のデコード処理を適用していることに注意してください。左画像のPSNRは、実験設定に関係なく同じで、38.38744に等しい。表7は、本適応手法と[11,13,14]の手法との比較である。また、標準MPO法[1]を適用した場合の性能も報告した。Compression Ratioの指標では、適応的な手法が他の手法を大きく上回っていることが観察できる。PSNRについては、[14]で提案されたハフマン符号化が最も良い結果を得ていますが、圧縮率(CR)の観点からは非常に低い性能となっています。

表6.LIVE 3Dデータセット[34]に適用した[20]メソッドの性能。

	ロッキー (dB)	ビットレート 保存	時間 (s)	シーア ール	PSNR 右
LmedSHarris70	1.4893	34.1709	2.0620	14.1925	36.6341
LmedSIFT70	1.4690	34.1709	1.6780	14.1925	36.6544
Norm8PointsHarris70	1.6734	34.1709	3.9440	14.1925	36.4500
ノーム8ポイント SIFT70	1.4604	34.1709	1.5960	14.1925	36.6631
LmedSHarris65	1.6804	40.3686	1.1460	14.7500	36.4430
LmedSIFT65	1.6025	40.3686	4.3960	14.7500	36.5209
Norm8PointsHarris65	1.7414	40.3686	3.4940	14.7500	36.3820
ノーム8ポイント SIFT65	1.6379	40.3686	1.0060	14.7500	36.4856

表7. LIVE 3Dデータセット[34]における評価手法の圧縮率 (CR) と平均PSNR。

方法	シーア ール	PSNR
アダプティブ [20]	14.4713	37.4583
標準MPO	11.6094	38.2526
ロッキーSICCAC[11] の場合	8.6400	41.5831
ロッキーS2ICAC [13] (英語 SICHC [14])	6.3600	33.9400
ロスレスSICCAC[11] の場合	4.4271	49.1446
ロスレスS2ICAC[13] の場合	4.1488	41.7359
	3.6100	34.885

4.2. 主観評価

CR の点では適応的手法が最も優れているが、適応的手法で得られた再構成後の画像の知覚品質を評価するために、本項で述べる主観的品質評価実験を設定した。上記で報告した適応法のビットレート節約量、PSNR、計算時間などの品質評価に加え、再構成画像の品質に関する主観評価も設定した。実験には、性別、年齢、視力に問題がある人などのばらつきを考慮して選んだ116人が参加した。具体的には、各

参加者は、矯正レンズまたは眼鏡を使用しているかどうか、性別、年齢を、年齢範囲（18-29、30-39、40-49、50-59、60歳以上）から選択した。50-59 歳は 10 名と少数であったため、最後の 2 つの年齢層を統合し、50 歳以上という新たなカテゴリーを作成した。1 枚はオリジナルの画質のステレオペアの右側の画像、もう 1 枚は本論文で述べる圧縮・再構成パイプライン後の同じ画像の画質を評価するよう各ユーザに要求した。各ユーザはMiddleburyデータセットからランダムに選んだ10組の画像を評価し、低品質の画像の位置（左または右）もランダムに選択しました。ユーザが低画質画像を正しく選択した回数（ヒット）と、それ以外を選択した回数（ミス）をカウントしました。表8は、異なるユーザーカテゴリーを考慮したヒットとミスの割合を示している。また、図8に示す円グラフでも、一般的なヒット・ミスの割合と詳細なヒット・ミスの割合を観察することができる。この結果から、参加者は元の画像と再構成された画像を区別することができないことがわかる。この結果は、性別、年齢、視力障害の有無によるマージンも考慮しても不変である。また、独立カイ2検定も行い、考慮したカテゴリーと正しい画像を推測する能力との間の独立性を確認した。この検定は、いくつかの有意な値(1)から(3)までを考慮して繰り返された。

0.01~0.05) では、帰無仮説 H_0 (すなわち独立) が、常に99%以上の検定信頼度で受け入れられました。

表8. 主観評価実験で収集されたヒットとミス。

	ヒット	ミス
女性	0.537931	0.462069
男性	0.593620	0.406380
メガネなし	0.594990	0.405010
メガネをかける と	0.562030	0.437970
18歳~29歳	0.613269	0.386731
30歳~39歳	0.514563	0.485437
40-49歳	0.581818	0.418182
年齢>50歳	0.482759	0.517241

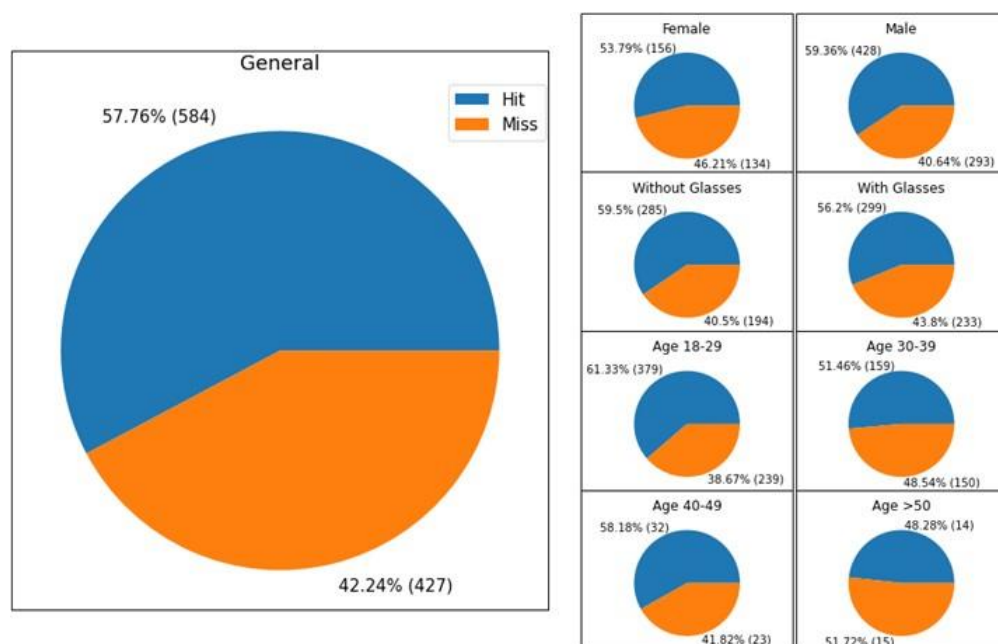


図8.主観評価で得られたヒット／ミスの詳細性能。

5. 結論と今後の課題

本論文では、まず、立体画像圧縮の概要について、圧縮画像の標準化の問題に焦点をあてて紹介する。次に、[19]で発表され、[20]でさらに拡張された適応的立体視画像圧縮のアプローチについて調査し、詳細を説明する。特に、解析対象であるアプローチのマッチングストラテジーの進歩が示されている。次に、適応的ステレオ圧縮法について、圧縮画像品質、2種類の最適化手法、2種類のキーポイント抽出手法を考慮した様々な設定での評価を行う。得られた16の評価設定を用いて、2001年から2014年の間に公開された5つのMiddleburyデータセットで圧縮法のベンチマークを行ったが、これは、ステレオスコーピー分野におけるアルゴリズムの客観的評価のための確立された基準を示すものである。その結果、本方式は、視覚的品質を維持したまま高圧縮された立体視画像を得ることができ、追加の保存ペイロードを必要とせず、[1]仕様に準拠したファイル形式を使用することができることが示されました。本ベンチマークでは、多数の参加者が提案された2つの画像の中から最高品質の画像を選択することで、再構成画像の知覚品質も評価しました。本手法の将来の拡張として、より高度で特殊な人間の視覚システムに基づく品質メトリック[35]を考慮して性能を評価し、シーンの3次元知覚の品質を高く維持しながら低画質の画像をどの程度圧縮できるかを適切に評価することが可能です。画像の品質を大幅に下げる拡張実験を行い、2の累乗（8 8、16 16、32 32など）の解像度次元の画像ブロックを使用した場合の効果（ブロッキングアーテファクトなど）を観察することも可能です。さらに、立体視画像圧縮の研究は、立体視が適用されている新しいアプリケーションやドメインについても行うことができます[36]。

XXXX
補足資料です。以下は、<https://www.mdpi.com/article/10.3390/jimaging7080160/s1> でオンライン公開されています。
.3390/jimaging7080160/s1, 表S1 : Middlebury-scenes 2014データセットでの結果、低品質65を考慮、表S2 : Middlebury-scenes 2014データセットでの結果、低品質65を考慮。Middlebury-scenes 2014データセットでの結果、70の低品質を考慮、表S3:Middlebury-scenes 2006データセットでの結果、低品質65を考慮、表S4.Middlebury-scenes 2006データセットでの結果、低品質70を考慮。Middlebury-scenes 2006データセットでの結果、低品質70を考慮、表S5: Middlebury- Scenes 2005データセットでの結果、低品質65を考慮、表S6: Middlebury-scenes 2006データセットでの結果、低品質70を考慮、表S6: Middlebury-scenes 2006データセットでの結果、低品質65を考慮 Middleburyscenes 2005 データセットでの結果（低品質 70 を考慮）、表S7. 表S7 : Middlebury-scenes 2003データセットに関する結果、65の低品質を考慮、表S8 : Middlebury-scenes 2003データセットに関する結果、70の低品質を考慮、表S9 : Middlebury-scenes 2001データセットに関する結果、65の低品質を考慮、表S10 : Middlebury-scenes 2001データセットに関する結果、70の低品質を考慮。

著者の貢献概念化、A.O., S.B. and F.R., 方法論、A.O., S.B. and F.R., ソフトウェア、A.O., 検証、A.O., 形式分析、A.O., S.B. and F.R., 調査、A.O., S.B. and F.R., データキュレーション、A.O., A.O., S.B. and F.R., Data Curation, M.G.執筆-原案作成、A.O., 執筆-レビューおよび編集、A.O., S.B., F.R., 可視化、A.O., M.G., 監督、S.B., F.R. 著者全員が本原稿を読み、合意している。

資金提供この研究は、外部からの資金援助を受けていない。施

設審査委員会の声明該当なし。インフォームド・コンセント（同意書）。該当なし。

参考文献

データの利用可能性に関する声明この論文で使用されたデータは公開されています。

利益相反。著者は利益相反を宣言していない。

1. マルチピクチャーフォーマット-ホワイトペーパー DC-X007-2009.Available online: https://www.cipa.jp/std/documents/e/DC-X007-KEY_E.pdf (accessed on 22 August 2021).
2. Wallace, G.K. 静止画圧縮規格JPEG. *IEEE Trans. Consum. Electron.* **1992**, *38*, 18-34.[[クロスレフ](#)].
3. Hudson, G.; Léger, A.; Niss, B.; Sebestyén, I.; Vaaben, J. JPEG-1 standard 25 years:過去、現在、そして未来の成功の理由。 *J. Electron. Imaging* **2018**, *27*, 040901.[[CrossRef](#)] ([クロスレフ](#))

4. Battiato, S.; Mancuso, M.; Bosco, A.; Guarnera, M. Psychovisual and statistical optimization of quantization tables for DCT compression engines. (DCT圧縮エンジンの量子化テーブルの心理的・統計的最適化).第11回画像解析・処理国際会議論文集, イタリア、パレルモ, 2001年9月26-28日; pp.602-606.
5. Christopoulos, C.; Skodras, A.; Ebrahimi, T. JPEG2000 静止画符号化方式. 概要*IEEE Trans.Consum.Electron.* **2000**, *46*, 1103-1127.[参照] [\[CrossRef\]](#)
6. H.264ベースの立体視映像符号化へのアプローチ.H.264ベースの立体映像符号化のアプローチ. In *Proceedings of the Third International Conference on Image and Graphics (ICIG'04)*, Hong Kong, China, 18-20 December 2004; pp.365-368.
7. Woo, W.; Ortega, A. MRFモデルを用いた視差補正によるステレオ画像圧縮.視覚通信と画像処理'96の議事録. *International Society for Optics and Photonics*, Orlando, FL, USA, 17-20 March 1996; Volume 2727, pp.28-42.にて。
8. このような場合、「ステレオ画像符号化」のためのFFTベースの視差推定を行う。In *Proceedings of the 2003 International Conference on Image Processing (Cat. No. 03CH37429)*, Barcelona, Spain, 14-17 September 2003; Volume 1, pp.1-761.
9. Schenkel, M.B.; Luo, C.; Frossard, P.; Wu, F. Joint decoding of stereo JPEG image Pairs (ステレオJPEG画像ペアの共同デコード).2010 IEEE International Conference on Image Processing, Hong Kong, China, 26-29 September 2010; pp.2633-2636 に収録されています。
10. Kadaikar, A.; Dauphin, G.; Mokraoui, A. Joint disparity and variable size-block optimization algorithm for stereoscopic image compression (立体視画像の視差と可変サイズブロックの最適化アルゴリズム).*Signal Process.Image Commun.***2018**, *61*, 1-8.[CrossRef] (英語)
11. Poolakkachalil, T.; Chandran, S. カーブレットを用いた立体視画像圧縮.*J. Theor.Appl.Inf.Technol.***2019**, *97*, 1146-1154.
12. Kumari, P.D.; Panigrahi, S.K.; Narayana, M. *Image Compression Algorithm Based on Curvelet Transforms and Comparative Analysis with JPEG and JPEG 2000*; International Journal of Electronics Engineering Research; Research India Publications:Delhi, India, 2017; Volume 9, pp.1373-1387.
13. Poolakkachalil, T.K.; Chandran, S. 算術符号化を用いた和算的立体視画像圧縮.インドネシア. *J. Electr.Eng.インフォマティクス IJEEI* **2019**, *7*, 564-576.
14. Poolakkachalil, T.K.; Chandran, S. Analysis of Stereoscopic Image Compression Using Arithmetic Coding and Huffman Coding (算術符号化とハフマン符号化を用いた立体画像圧縮の分析).2018 International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, 11-12 July 2018; pp.214-220 に掲載されています。
15. Deng, X.; Yang, W.; Yang, R.; Xu, M.; Liu, E.; Feng, Q.; Timofte, R. Deep Homography for Efficient Stereo Image Compression (効率的なステレオ画像圧縮のための深層ホモグラフィ). In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Virtual Conference*, 19-25 June 2021; pp.1492-1501.
16. Bao, W.; Wang, W.; Xu, Y.; Guo, Y.; Hong, S.; Zhang, X. InStereo2K.屋内シーンのステレオマッチングのための大規模な実データセット. InStereo2K: A large real dataset for stereo matching in indoor scenes.屋内シーンにおけるステレオマッチングのための大規模な実データセット. *Sci. China Inf.Sci.* **2020**, *63*, 1-11.[CrossRef] (英語)
17. Geiger, A.; Lenz, P.; Stiller, C.; Urtasun, R. Vision meets robotics (視覚とロボット工学の出会い)。kittiデータセット.を使用。*J. Robot.Res.* **2013**, *32*, 1231-1237.[CrossRef] (英語)
18. Kadri, I.; Dauphin, G.; Mokraoui, A.; Lachiri, Z. Stereoscopic image coding using a global disparity estimation algorithm optimizing-ing the compensation scheme impact.2020年信号処理学会講演論文集:Algorithms, Architectures, Arrangements, and Applications (SPA), Poznan, Poland, 23-25 September 2020; pp.69-74.
19. Ortis, A.; Rundo, F.; Di Giore, G.; Battiato, S. Adaptive Compression of Stereoscopic Images (立体画像の適応的圧縮). In *Image Analysis and Processing- ICIAP 2013*; Petrosino, A., Ed.; Springer: Berlin/Heidelberg, Germany, 2013; pp.391-399.にて。
20. Ortis, A.; Battiato, S. A new fast matching method for adaptive compression of stereoscopic images (立体画像の適応的圧縮のための新しい高速マッチング法)。三次元画像 処理、計測 (3DIPM)、および応用 **2015**.*Int.Soc. Opt.Photonics* **2015**, *9393*, 93930K.
21. O'Mahony, N.; Campbell, S.; Carvalho, A.; Harapanahalli, S.; Hernandez, G.V.; Krpalkova, L.; Riordan, D.; Walsh, J. Deep learning vs. traditional computer vision (深層学習と従来のコンピュータビジョン)。科学と情報会議,シュブリンガーにて。ラスベガス、ネバダ州、アメリカ、2019年; pp.128-144.
22. Briechele, K.; Hanebeck, U.D. 高速正規化相互相関を用いたテンプレートマッチング。光学的パターン認識 **XII**.*Int.Soc. Opt.Photonics* **2001**, *4387*, 95-102.
23. Hartley, R.; Zisserman, A. *Multiple View Geometry in Computer Vision*; Cambridge University Press:Cambridge, UK, 2003.
24. Scharstein, D.; Szeliski, R. High-accuracy stereo depth maps using structured light.In *Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Madison, WI, USA, 18-20 June 2003; Volume 1, p. 1.
25. Longuet-Higgins, H.C. A computer algorithm for reconstructing a scene from two projections (2つの投影からシーンを再構成するコンピュータ・アルゴリズム)。ネイチャー **1981**, *293*, 133.[参照] [\[CrossRef\]](#)
26. Lowe, D.G. スケール不変のキーポイントからの特徴的な画像特徴. *Int.J. Comput.Vis.***2004**, *60*, 91-110.[CrossRef] (英語)
27. Harris, C.G.; Stephens, M. A combined corner and edge detector (コーナーとエッジの複合検出器).*Alvey Vision Conference; Citeseer:Manchester, UK*, 1988; Volume 15, pp.10-5244.
28. 自己組織化マップ.*IEEE* **1990**, *78*, 1464-1480.[クロスリード]
29. 3Dメディア-3Dテクノロジーとソフトウェア.2013.Available online: <http://www.3dmedia.com/gallery> (accessed on 2 July 2021).
30. Scharstein, D.; Szeliski, R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms (密な2フレームステレオ対応アルゴリズムの分類と評価). *Int.J. Comput.Vis.* **2002**, *47*, 7-42.[クロスリード] [\[CrossRef\]](#)

31. Scharstein, D.; Pal, C. Learning conditional random fields for stereo. In Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, USA, 17-22 June 2007; pp.1-8.
32. Hirschmuller, H.; Scharstein, D. Evaluation of Cost Functions for Stereo Matching (ステレオマッチングのためのコスト関数の評価). 2007 IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, USA, 17-22 June 2007; pp.1-8.

33. Scharstein, D.; Hirschmüller, H.; Kitajima, Y.; Krathwohl, G.; Nešić, N.; Wang X.; Westling, P. *High-Resolution Stereo Datasets with Subpixel-Accurate Ground Truth*; Lecture Notes in Computer Science; Springer International Publishing: Munster, Germany, 2014; pp.31-42を参照。
34. Moorthy, A.K.; Su, C.C.; Mittal, A.; Bovik, A.C. 立体視映像の主観的画質評価。 *信号処理.画像通信* **2013**, 28, 870-883.[[クロスリード](#)].
35. Banitalebi-Dehkordi, A.; Pourazad, M.T.; Nasiopoulos, P. A human visual system-based 3D video quality metric.これは、人間の視覚系に基づいた3D映像の品質指標である。In *Proceedings of the 2012 International Conference on 3D Imaging (IC3D)*, Liege, Belgium, 3-5 December 2012; pp.1-5.
36. Duarte, J; Cassin, R; Huijts, J; Iwan, B; Fortuna, F; Delbecq, L; Chapman, H; Fajardo, M; Kovacev, M; Boutu, W; et al. コンピュータステレオレンズレスX線イメージング. *Nat.Photonics* **2019**, 13, 449-453.[[CrossRef](#)] (英語)