# セマンティック通信を用いたリアルタイム映像伝送の研究

研究代表者　　伊藤 暢彦　　日本工業大学 先進工学部 准教授

## 1 研究課題の概要

本研究課題では「セマンティック通信を用いた低コストかつリアルタイムな映像伝送の実現」を目的として，意味情報に基づく生成型動画伝送手法，意味情報の抽出手法，及び実機実装による伝送システムの設計・評価に関する研究を推進した. 本報告書では, 研究期間中に実施した各検討内容とその成果について述べる.

## 2 研究成果の報告

本研究期間中に取り組んだ研究内容を以下に示す.

- **意味情報に基づく生成型動画伝送手法**・・・セマンティック通信を用いたリアルタイムな映像伝送を実現するにあたり，予め撮影された動画を伝送する際の送信データ量を削減するための手法を検討し，提案手法による送信データ量の削減効果や復元された動画の再現性の検証を実施した.
- **シーングラフを用いた意味情報の抽出手法**・・・セマンティック通信を用いた映像伝送では，映像に含まれる特定の意味情報に着目した低コストな映像伝送を実現するため，受信端末側の要求に応じて抽出する意味情報を柔軟に選択できる必要がある．そこで，シーングラフを活用することによって，受信端末側の要求に応じて送信端末が意味情報を柔軟に取捨選択できる意味情報の抽出方法について検討を行った.
- **狭帯域通信技術を用いた生成型画像伝送システムの実装**・・・既存研究や上述した検討内容では，実機実装を踏まえたシステム全体の設計について未検討であった．そこで，実機を用いた映像伝送の実現に向けて，狭帯域通信技術を用いた画像伝送システムについて検討し，LoRa を用いたエンドエンド間での画像伝送実験を行った.

各研究における詳細な成果については次節以降に述べる.

### 2-1 意味情報に基づく生成型動画伝送手法

(1) はじめに

本研究では，セマンティック通信によって動画伝送に係る通信コストを削減するための手法として，意味情報に基づく生成型動画伝送手法を提案し，その基礎的な性能を評価した.

画像伝送に係る送信データ量を削減するアプローチとして，画像に含まれる特定の意味情報に基づき画像伝送を実現する生成型画像伝送手法が提案されている[1]．本手法では，送信端末側が，画像に含まれる複数種類の意味情報を機械学習モデルによって抽出し，抽出された意味情報のみを伝送することによって，一般的な画像圧縮アルゴリズムを用いた画像伝送に比べて送信データ量を大幅に削減する．その後，受信端末は生成モデルを用いて受信した意味情報と同一の意味情報を有する画像を復元することにより，エンドエンド間での画像伝送を実現する．なお，本手法を含む，データの意味に着目した通信技術はセマンティック通信と呼ばれ，ビット列に基づく既存の通信に代わる新たな技術として注目されている[2]～[4].

本手法は，画像に含まれる特定の意味情報に基づいた画像伝送を実現するため，画像内に映る特定のオブジェクトを対象とした監視や物体検知等，画像内のある意味情報が重要視されるユースケースにおいて有効であると考えられる．そして，本手法を動画伝送へ応用することができれば，動画伝送に係る伝送コストを大幅に削減できる可能性があり，大規模かつ低コストな映像監視などの実現に寄与できると期待される．しかし，本手法を動画伝送へ直接適用した場合，動画を構成する全てのフレームに対して意味情報の抽出を行う必要があり，送信端末における意味情報抽出に伴う計算コストや処理遅延の増大を引き起こす可能性がある．また，動画内の連続するフレーム間には時間的な相関が存在するため，全てのフレームから個別に意味情報を抽出し伝送することは，伝送効率の観点から非効率であると考えられる.

そこで本研究では，フレーム補間技術を用いることで，意味情報の抽出及び伝送に係る計算コストや送信データ量を削減しながら動画伝送を実現する生成型動画伝送手法を提案した.



公益財団法人電気通信普及財団
研究調査助成報告書 第41号 2026年度

(2) 関連研究

意味情報に基づく生成型画像伝送手法[1]では，伝送対象である画像に含まれる特定の意味情報に着目した画像伝送を実現する．本手法における送信端末及び受信端末の動作手順を以下に示す．

**送信端末による意味情報の抽出及び伝送：**送信端末はまず画像キャプショニングモデルを用いて，元画像全体の様子を自然言語によって説明するキャプションを作成することによって，元画像から説明情報を抽出する．この説明情報には，元画像に写るオブジェクトの種類や大まかな様子，周辺の背景に関する情報など，自然言語によって説明可能な意味情報が含まれている．これと同時に送信端末は，セマンティックセグメンテーションを用いて元画像の各ピクセルに対してラベル付けを行い，セグメント配列を作成することで，元画像から領域情報を抽出する．この領域情報には，元画像に写るオブジェクトの位置や構図に関する意味情報が含まれている．さらに，送信端末は，セグメント配列に含まれる各ラベルに対応するピクセルの RGB 平均値を算出し，カラーパレットを作成することによって，元画像の各領域に対する色情報を抽出する．以上の手順によって送信端末は，元画像がもつ複数種類の意味情報を抽出し，これらの情報のみを伝送することによって，画像伝送に係る伝送コストを削減する．なお，先行研究における送信データ量の測定結果では，上述した意味情報の合計データ量は，JPEG 形式の元画像のデータ量と比較して約 1/14 であることが示されている[1]．

**受信端末による画像の復元：**前述した手順によって抽出された意味情報を受信すると，受信端末は受信した情報を画像生成モデルへ入力することによって画像を復元する．なお，このとき，受信端末で使用する画像生成モデルは，一般的に入力されたテキスト及び画像に基づいて画像生成を行う．そのため受信端末は，セグメント配列内の各ラベルをカラーパレット内の対応する RGB 値に置換することによって，領域分割画像を作成する．その後，受信端末はキャプションと領域分割画像を画像生成モデルへ入力することによって，元画像と同等の意味情報を有する画像を復元する．以上の手順を通じて，本手法では意味情報に基づく低コストな画像伝送を実現する．

上述した手法は，画像に含まれる特定の意味情報に着目することで，画像伝送に係る送信データ量を大幅に削減する．そして，本手法を動画伝送へ応用することができれば，動画伝送に係る通信帯域の消費量を大幅に削減できる可能性があり，帯域制限の厳しい IoT 環境における大規模かつリアルタイムな映像監視などの実現に寄与できると期待される．しかし，本手法を動画伝送へ直接適用した場合，動画の全フレームに対して意味情報を抽出する必要があり，これは送信端末における意味情報抽出に伴う計算コストや処理遅延の増大を引き起こす可能性がある．また，動画内の連続するフレーム間には時間的な相関が存在するため，全てのフレームから個別に意味情報を抽出及び伝送することは，通信リソースの観点からも非効率である．

(3) 提案手法

前節にて述べた課題に対し，本研究では，フレーム補間モデルを用いた意味情報に基づく生成型動画伝送手法を提案する．提案手法では，連続するフレームの間に中間フレームを生成及び追加することが可能なフレーム補間モデルを用いることによって，意味情報の抽出頻度を抑制しながら意味情報に基づいたエンドエンド間での動画伝送を実現する．

提案手法の概要を図 1 に示す．まず送信端末は，伝送対象となる元動画をフレームに分割し，各フレームから意味情報の抽出を行う．このとき，送信端末が元動画の全フレームに対して意味情報の抽出を行うと，

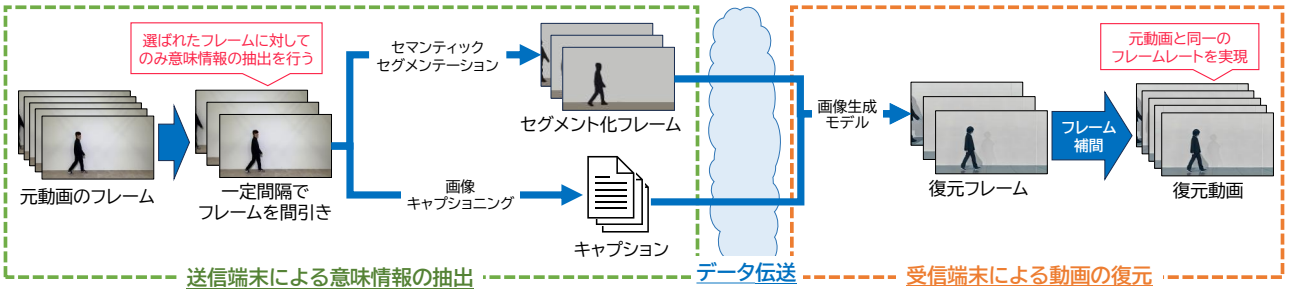


**図 1　意味情報に基づく生成型動画伝送手法の概要**

送信端末側の計算コストや送信データ量が増大する可能性がある．そこで，送信端末は全てのフレームから一定の間隔でフレームを選択し，選択されたフレームに対してのみ意味情報の抽出を行うことによって，意味情報の抽出に係る計算コスト及び送信データ量を削減する．そして，受信端末側で間引かれたフレームを補間することによって，元動画と同じフレームレートの動画の伝送を実現する（補間方法については後述する）．例えば，受信端末によるフレーム補間枚数を 1 枚に設定した場合，送信端末は 2 フレームに 1 回の頻度で意味情報の抽出を行う．送信端末は，従来手法と同様に，画像キャプショニングモデルを用いて抽出対象となるフレームから説明情報を抽出する．それと同時に，セマンティックセグメンテーションモデルを用いてオブジェクトの位置及び構図情報を抽出し，セグメント化されたフレームを作成する．その後，送信端末は抽出された意味情報のみを伝送する．

その後，受信端末は，これらの意味情報を受信すると，従来手法と同様に受信した意味情報を画像生成モデルへ入力することにより，フレームを復元する．その後，受信端末は連続するフレーム間の中間フレームを生成することが可能なフレーム補間モデルを用いることによって，送信端末側で意味情報の抽出を省略されたフレームを補間し，元動画と同一のフレームレートを有する動画を復元する．

(4) 性能評価

動画伝送に係る送信データ量，及び提案手法によって復元された動画の視覚的な再現性の観点から，提案手法の基礎的な性能を評価する．

**実験環境**

本実験では，図 2(a)に示す，1 名の男性がフレームの左端から右端へ歩行する様子を写した MP4 形式の動画を対象とする．実験に使用した動画の詳細な情報は表 1 に示す．本実験ではまず，この動画を 407 枚のフレームに分割し，各フレームに対して画像キャプショニングモデル及びセマンティックセグメンテーションモデルを用いてキャプションとセグメント化されたフレームを作成する．なお，本実験では画像キャプショニングモデルには Florence-2[5]を，セマンティックセグメンテーションモデルには SegFormer[6]を使用した.その後,各フレームに対応するキャプションとセグメント化されたフレームを画像生成モデルへ入力し，フレームを復元する．ここで，本実験では画像生成モデルとして，高速な画像生成が可能な StreamDiffusion[7]を使用した．そして，フレーム補間枚数を 1 枚及び 3 枚に設定した場合において送信端末が伝送する意味情報の合計送信データ量と受信端末によって復元される動画の視覚的な特徴を比較評価する．なお，本実験ではフレーム補間モデルとして FILM[8]を使用した．

**実験結果**

元動画のデータ量（MP4 形式）と提案手法によって伝送される意味情報のデータ量の比較結果を表 2 に示す．結果より，提案手法によって抽出された意味情報は，MP4 形式の元動画と比べてデータ量を大幅に削減できていることが分かる．これは，意味情報の抽出によるデータ量の圧縮に加え，意味情報抽出及び伝送の対象とするフレームを全フレームの中から選択しているためである．以上の結果より，送信端末側による意味情報の抽出及び受信端末側によるフレーム補間を組み合わせた動画伝送は，動画伝送に係る伝送コスト削減に有効なアプローチであることが判明した．

**表 1　実験に使用した元動画の詳細情報**

| | |
|---|---|
| ファイル形式 | MP4 |
| 解像度 | 640 × 360 ピクセル |
| フレーム数 | 407 |
| フレームレート | 60 fps |
| 動画時間 | 6.78 秒 |

表 2　送信データ量の比較結果

| 元動画（MP4 形式） | 提案手法<br>(フレーム補間枚数：1 枚) | 提案手法<br>(フレーム補間枚数：3 枚) |
|---|---|---|
| 793.662 KB | 462.825 KB | 234.711 KB |

(a) 元動画

(b) 提案手法(フレーム補間枚数:1枚)

(c) 提案手法(フレーム補間枚数:3枚)

図 2　元動画及び提案手法によって復元された動画のフレーム（抜粋）

元動画及び提案手法によって復元された動画のフレームの一部を図 2 に示す．結果より，提案手法によって復元された動画は，元動画における「一人の男性が，白い壁の前をフレームの左端から右端に向かって歩いている」という元動画もつ本質的な意味情報を適切に再現できていることが分かる．一方，動画内に映る人物の顔や服装のような，元動画がもつより詳細な意味情報は抽象化されていることが分かる．これは，送信端末による意味情報の抽出によって，元動画がもつ一部の意味情報が消失しているためである．このことから，提案手法を含むセマンティック通信では，受信端末側の要求やユースケース等に応じて抽出する意味情報の種類や粒度を適切に調節することが重要であることが確認された．

### 2-2 シーングラフを用いた意味情報の抽出手法

(1) はじめに

本研究では，前述した意味情報に基づく生成型画像伝送及び生成型動画伝送における意味情報の新たな抽出方法として，シーングラフを活用する手法を提案した．

前述した意味情報に基づく画像及び動画伝送手法では，送信端末が画像キャプショニングモデル及びセマンティックセグメンテーションを用いて，画像や動画に含まれる意味情報の抽出を行う．これらの意味情報の抽出手法は，メディアがもつ意味情報を可能な限り保持しながら伝送を行うが，これらの意味情報が受信端末側の要求する情報と完全に一致するとは限らず，伝送された意味情報のうち一部の情報のみで満足する可能性が考えられる．そのため，受信端末側の要求に応じて送信端末側が抽出及び伝送する意味情報を取捨選択する必要が考えられるが，その実現性については十分に検討がなされていない．

これに対し，本研究ではメディア内に映るオブジェクトの構成や関係性をグラフ構造で表現するシーングラフを用いることで，受信端末側の要求に応じて伝送対象とする意味情報の取捨選択を可能とする意味情報抽出手法について提案を行った．

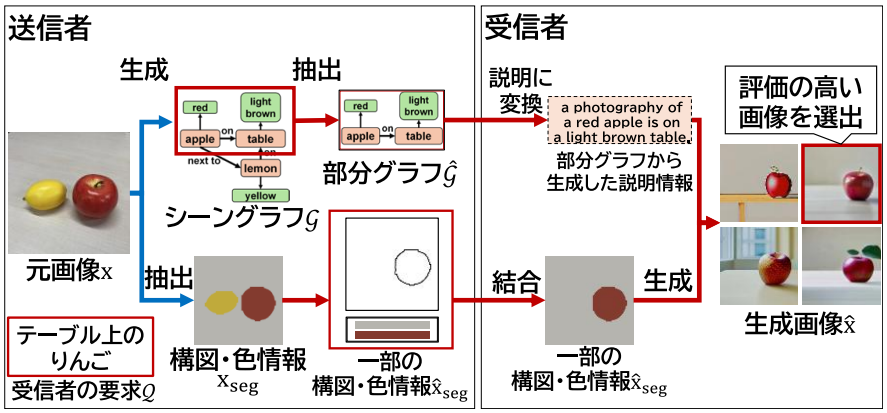


図 3　提案手法の動作例

(2) 提案手法

本研究では，シーングラフを活用することによって，受信端末側の要求に応じた意味情報の柔軟な抽出を行う手法を提案し，セマンティック通信の数理フレームワーク[9]を参考に提案手法の定式化を検討した．

提案手法の動作例を図 3 に示す．まず，送信者は，伝送対象となる画像$x$から，被写体集合$\mathcal{V} = \{1,2,\dots,V\}$と被写体$v, v' \epsilon \mathcal{V}$間の関係$(v, v')$からなる関係集合$\{(v, v') | v, v' \in \mathcal{V}, v \neq v'\}$によって構成される自然言語で表現されたシーングラフ$\mathcal{G} = (\mathcal{V}, \mathcal{G})$を生成する．シーングラフにおいて，被写体はその形状，色など視覚的な表現が容易である一方，被写体間関係は相対的な状態や動作などを表すため，近い・遠いなど，意味的な曖昧性を含む可能性がある．そこで，被写体間の関係を空間的に補正するため，元画像$x$に対してセグメンテーションを行い，各被写体の輪郭と平均色を保持するセグメンテーション画像$x_{seg}$を得る．

次に，送信者は，被写体に対する要求$\mathcal{Q}_{\mathrm{obj}}$と被写体間関係に対する要求$\mathcal{Q}_{\mathrm{rel}}$の組で構成される，受信者の要求$\mathcal{Q} = (\mathcal{Q}_{\mathrm{obj}}, \mathcal{Q}_{\mathrm{rel}})$に基づき，送信する意味情報を選択する．ここで，各被写体と被写体間の各関係はそれぞれが様々な属性からなる集合と捉え，ある側面から意味的に分類するための基準$k$に基づき，分類を与えるカテゴリ$\mathcal{C}_k$を考える．これにより，カテゴリ$\mathcal{C}_k$を用いて各被写体や被写体間の各関係をある観点から分類することで，それぞれがもつ属性が得られ，分類不可能な場合，その属性が得られないものとする．ここで，受信者が要求する被写体となり得る候補の集合$\mathcal{U}$を考え，各被写体候補に対する要求$\mathcal{Q}_{\mathrm{obj}} = \{\mathcal{K}^{(u)} \mid u \in \mathcal{U}\}$は，各候補$u$に対する分類基準集合$\mathcal{K}^{(u)}$を要素とする集合族とし，被写体候補間の各関係に対する要求$\mathcal{Q}_{\mathrm{rel}} = \{\mathcal{K}^{(u,u')} \mid u, u' \in \mathcal{U}, u \neq u'\}$は，異なる 2 つの被写体候補$u - u'$間の関係に対する分類基準集合$\mathcal{K}^{(u,u')}$を要素とする集合族とする．つまり，$\mathcal{K}^{(u)}$と$\mathcal{K}^{(u,u')}$は，受信者が要求する個々の被写体やその関係に対し，複数の分類基準を同時に満たす意味的要求となる．

したがって，シーングラフ$\mathcal{G}$の各被写体$v \in \mathcal{V}$と各関係$(v, v') \in \mathcal{E}$に対し，各候補$u$に対する分類基準集合$\mathcal{K}^{(u)} \in \mathcal{Q}_{\mathrm{obj}}$の各基準$k^{(u)} \in \mathcal{K}^{(u)}$によるカテゴリ$\mathcal{C}_{k^{(u)}}$を，異なる候補$u - u'$間の関係に対する分類基準集合$\mathcal{K}^{(u,u')} \in \mathcal{Q}_{\mathrm{rel}}$の各基準$k^{(u,u')} \in \mathcal{K}^{(u,u')}$による$\mathcal{C}_{k^{(u,u')}}$それぞれから要求に基づく属性集合を抽出し，各被写体$\mathcal{V}$の要素$\hat{v}, \hat{v}'$とその各関係$(\hat{v}, \hat{v}') \in \hat{\mathrm{E}}$を得る．

属性が得られなかった各被写体$v$と各関係$(v, v')$, 2 つの被写体$\hat{v}, \hat{v}'$のいずれかが存在しない$(\hat{v}, \hat{v}')$を$\mathcal{V}$と$\hat{\mathrm{E}}$から除外し，部分グラフ$\mathcal{G} = (\mathcal{V}, \hat{\mathrm{E}})$を構成する．$x_{\mathrm{seg}}$に対しても要求に基づく抽出を行い，一部の構図情報と色情報からなるセグメンテーション画像$\hat{x}_{\mathrm{seg}}$を作成し，部分グラフ$\mathcal{G}$とともに受信者へ送信する．

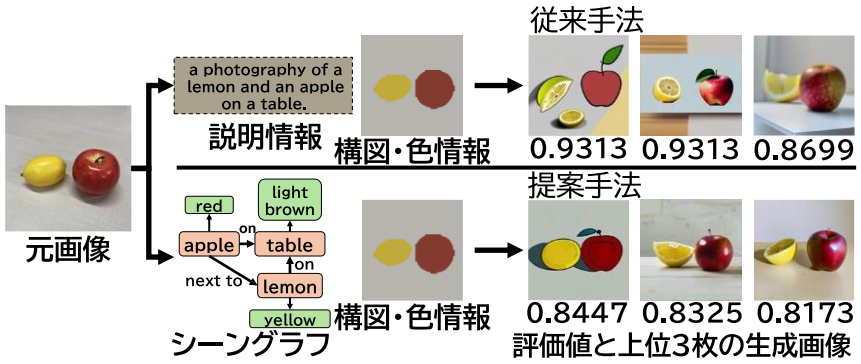


図 4　提案手法と既存手法の比較結果（評価値上位 3 枚）

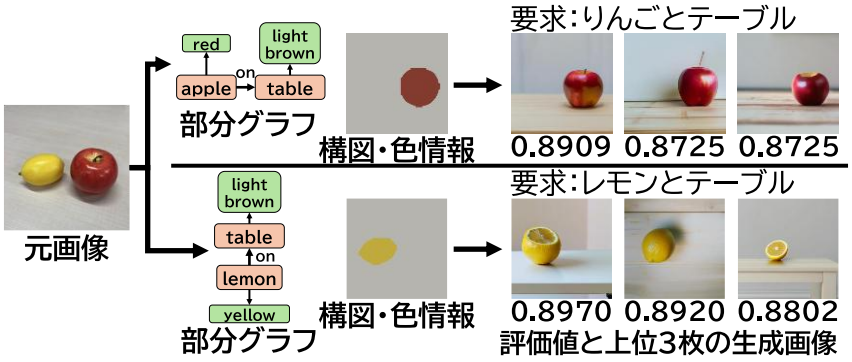


図 5　意味情報を選択した場合生成結果（評価値上位 3 枚）

受信者は，受信したシーングラフ$\mathcal{G}$から生成した説明情報と，セグメンテーション画像$\hat{x}_{\mathrm{seg}}$から，$N$枚の画像$X = \{\hat{x}_1, \ldots, \hat{x}_N\}$を生成する．次に，意味情報の類似度評価$S(\cdot,\cdot)$を用いて，受信した$\mathcal{G}$と各生成画像$\hat{x}_n$から抽出したシーングラフ$\mathcal{G}_n$との間の評価値$y_n = S(\mathcal{G}_n, \mathcal{G})$を算出し，その評価値集合$\mathcal{Y} = \{y_1, \ldots, y_N\}$の中で評価値が最大となる画像を伝送結果とする．

(3) 動作検証

提案手法の動作を確認するため，従来手法である意味情報の選択を行わない条件下での生成型画像伝送手法[1]と提案手法間の比較，及び提案手法における要求に基づく意味情報の選択とその結果を確認した．本実験では，$\hat{\mathcal{G}}_n$と$\mathcal{G}$間に基づく説明情報の意味的な類似度を，テキスト間類似度の評価指標である BERTScore[11]を用いて算出し，50 枚の生成画像のうちの上位 3 枚を比較した．

検証結果を図 4，5 に示す．図 4 より，意味情報の選択を行わない場合，従来手法と提案手法間では，生成画像に含まれる意味情報に大きな差異はなく，両者はおおよそ同程度の生成品質を維持できていると考えられる．そして，図 5 に示す通り，提案手法は元画像に含まれる意味情報のうち，受信側の要求によって指定された意味情報のみを含む画像が生成できていることが分かる．これは，部分グラフによる選択的な意味情報の伝送が可能であることを示している．

以上の結果より，シーングラフを活用することによって，受信端末側の要求に応じて伝送する意味情報の取捨選択が可能であることが確認された．なお，これらの検討では，受信端末側の要求を送信端末へフィードバックする手法が未検討である．そのため，現在，受信端末側の要求に応じて意味情報の抽出方法やデータ復元に使用する生成モデルの選択方法を制御することが可能な映像伝送システムを併せて検討している．

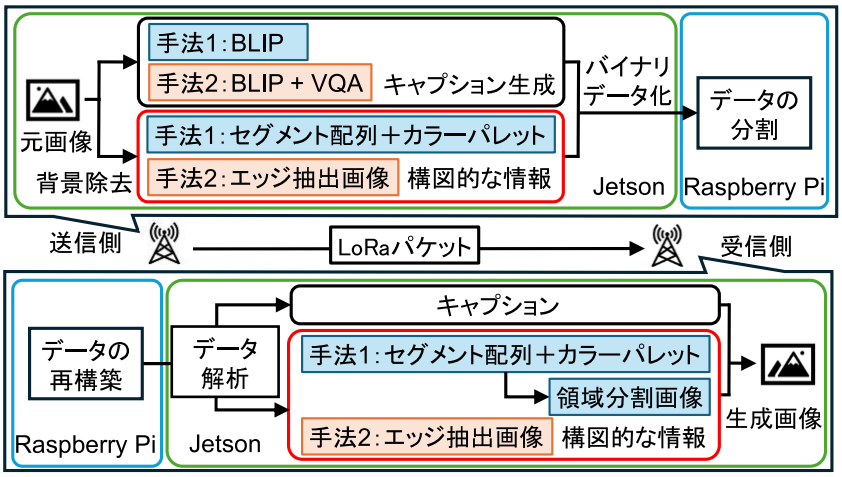


図 6　生成型画像伝送システムのシステムモデル

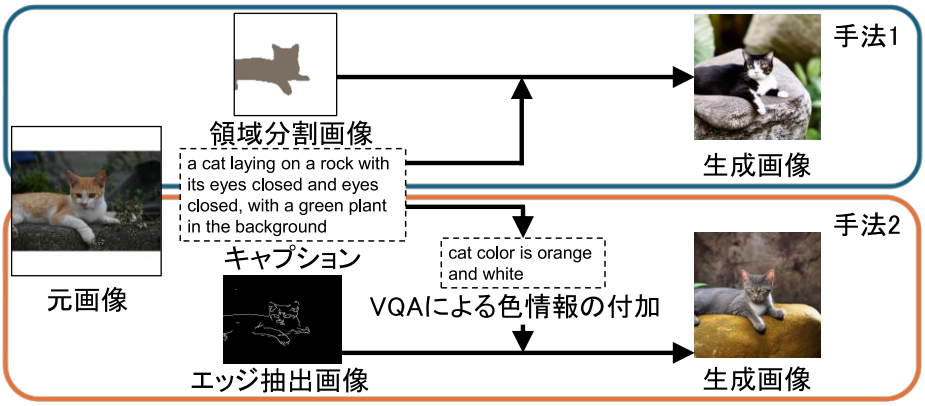


図 7　提案する画像伝送システムの動作例

表 3　通信実験の評価結果

| 手法 | 送信処理時間 | 通信時間 | 受信処理時間 | データ量 |
|---|---|---|---|---|
| 手法 1 | 9.3 秒 | 1.2 秒 | 13.2 秒 | 497.2 B |
| 手法 2 | 13.5 秒 | 3.6 秒 | 12.7 秒 | 1474.2 B |

### 2-3 狭帯域通信技術を用いた生成型画像伝送システムの実装

(1) はじめに

前述した意味情報に基づく生成型画像伝送手法及び動画伝送手法は，画像及び動画を伝送する際の送信データ量を大幅に削減することが可能である．しかし，これらの研究ではデータ量の比較評価や復元されたメディアの再現性に関する評価を行っているが，実機実装を踏まえたシステム全体の設計や狭帯域な通信環境下での実応用性については未検討となっている．これに対して本研究では，初期検討として既存手法である意味情報に基づく生成型画像伝送[1][10]を対象とし，狭帯域通信向けの意味情報に基づく生成型画像伝送システムの設計と実装を行い，実機実験を通じてその特性を解析した．

(2) 生成型画像伝送システムの実機実装

本研究で提案する生成型画像伝送システムのシステムモデルを図 6 に示す．ここで，既存手法である意味情報に基づく生成型画像伝送手法は，構図情報の抽出方法として前述したセグメント配列及びカラーパレットを用いるもの[1]（手法 1）と画像のエッジを抽出した画像を用いるもの[10]（手法 2）が提案されており，この両者についての設計と実装を検討する．本システムでは，送信側と受信側の両端末において，キャプション生成や画像生成等の AI 処理を担う Jetson AGX Orin と LPWA 通信規格の一つである LoRa による通信制御を担う Raspberry Pi 3B+で構成されており，両者は TCP によるソケット通信で連携する．

送信側の Jetson は撮影した画像を解析し，画像キャプショニングモデルの一種である BLIP を用いて自然言語による画像の説明文を抽出する．並行して，セマンティックセグメンテーションモデルの一種である DeepLabv3 により，伝送対象となる被写体ごとの領域を抽出し，受信端末側端末側で被写体の構図を適切に再現するため背景を除去する．ここで，手法 1 では，被写体ごとの RGB 平均値を保持するカラーパレットを作成する．手法 2 では，説明文抽出時に VQA (Visual question answering) で被写体の色を得るとともに，Canny 法でエッジ画像を抽出する．Jetson は，これらを含むバイナリデータとして結合し，Raspberry Pi へ転送する．Raspberry Pi は，受信したデータを LoRa の MTU に合わせて分割し，パケットとして順に送信する．

受信側の Raspberry Pi は受信したパケットからバイナリデータをバッファ上で再構築し，Jetson へ転送する．受信側の Jetson は受信したデータに基づき画像生成モデルにより画像を生成する．手法 1 では，セグメント配列とカラーパレットから作成した領域分割画像を，手法 2 ではエッジ画像を構図の制約として用い，自然言語による説明文をプロンプトとして画像を生成する．これにより，低帯域通信下においても送信元画像の被写体形状及び意味内容を保持した画像の再構築を実現する．

(3) 実機を用いた性能評価

提案システムにおける両手法の動作検証及び性能評価を行うため，ある 1 種類のオブジェクトが写った画像を伝送する実験を 5 枚の異なる画像に対して各 5 回，計 25 回実施した．本実験では，画像の解析開始からバイナリデータの転送完了までの時間 (送信処理時間)，先頭パケット送信開始から最後尾パケット受信完了までの時間 (通信時間)，受信側でのバイナリデータ受信から画像生成完了までの時間 (受信処理時間)，送信データ量をそれぞれ計測した．

提案システムの動作例を図 7 に，評価結果を表 3 に示す．元画像の平均データ量は 212.2 KB であり，表 3 より，提案手法いずれの場合においても送信データ量を削減できている．一方，手法 1 と手法 2 を比較した場合，送信処理にかかる時間，データ量共に手法 2 の方が大きい．これは，手法 1 では領域分割画像を送信するのに対し，手法 2 では被写体の輪郭に加えて内部のエッジ情報が含まれることからデータ量が増大したためであると考えられる．

## 3 まとめ

本研究では，「セマンティック通信を用いた低コストかつリアルタイムな映像伝送の実現」を目標として，研究期間中に以下の 3 つの課題に取り組んだ．

まず，意味情報に基づく生成型動画伝送手法については，フレーム補間モデルを組み合わせることで，全フレームに対する意味情報の抽出を省略しながらも，元動画と同一フレームレートの動画をエンドエンド間で伝送できることを示した．また実験により，提案手法によって伝送される意味情報の合計データ量は一般的な MP4 形式の動画と比較して大幅に削減されることが確認された．

シーングラフを用いた意味情報の抽出手法については，メディア内に映るオブジェクトの構成や関係性をグラフ構造で表現するシーングラフを活用することで，受信端末の要求に応じた選択的な意味情報の抽出・伝送が可能であることを動作検証によって確認した．また，従来手法と同等の生成品質を維持しながら特定の意味情報のみを含む画像を生成できることを併せて確認した．さらに，受信端末側の要求を送信端末側へフィードバックするための制御を加えた新たな映像伝送アーキテクチャについて検討を行った．

狭帯域通信技術を用いた生成型画像伝送システムの実装については，セマンティック通信による映像伝送システムの実機実装に向け，Jetson と Raspberry Pi 及び狭帯域通信技術の一つである LoRa を組み合わせた画像伝送システムを設計・実装し，端末間での画像伝送実験を通じてその動作特性を評価した．実験結果より，意味情報の抽出手法に依らず元画像と比べて送信データ量を削減できることが確認され，実機レベルでの低コストな画像伝送の実現可能性を示す成果が得られた．

今後の展望としては，まず，映像伝送におけるリアルタイム性の実現に向けて，意味情報の抽出，及び映像の復元に使用する各機械学習モデルの選定及び高速化について検討し，処理遅延に関する定量的な評価を行う必要がある．また，実機での伝送システムにおける伝送対象を画像から動画へと拡張し，LoRa 等の狭帯域通信環境下においても動画伝送が成立するかについて検証を行う必要がある．また，意味情報の選択手法については，受信端末の要求を送信端末へ動的にフィードバックするための機構を設計・実装し，要求駆動型の映像伝送システムとして統合する予定である．さらに，本研究期間中に実施した評価シナリオは全体を

通して限定的であったため，複数オブジェクトが登場するより複雑なシーンや，屋外・夜間等の多様な撮影条件下における頑健性の検証も今後実施する．これらの取り組みを通じて，IoT 環境における大規模かつリアルタイムな映像監視など，実社会への応用に向けた研究を継続・発展させる予定である．

## 【参考文献】


[1] E. Hosonuma, T. Yamazaki, T. Miyoshi, A. Taya, Y. Nishiyama, and K. Sezaki, "Image generative semantic communication with multi-modal similarity estimation for resource-limited networks," IEICE Trans. Commun., vol. E108-B, no. 3, pp. 260-273, March 2025.

[2] X. Luo, H.H. Chen, and Q. Guo, "Semantic communications: Overview, open issues, and future research directions," IEEE Wireless Commun., vol. 29, no. 1, pp. 210-219, Feb. 2022.

[3] D. Gunduz, Z. Qin, I.E. Aguerri, H.S. Dhillon, Z. Yang, A. Yener, K.K. Wong, and C.B. Chae, "Beyond transmitting bits: Context, semantics, and task-oriented communications," IEEE J. Selected Areas in Commun., vol. 41, no. 1, pp. 5-41, Jan. 2023.

[4] Z. Qin, X. Tao, J. Lu, W. Tong, and G.Y. Li, "Semantic communications: Principles and challenges," arXiv preprint arXiv:2201.01389, pp. 1-32, June 2021.

[5] B. Xiao, H. Wu, W. Xu, X. Dai, H. Hu, Y. Lu, M. Zeng, C. Liu, and L. Yuan, "Florence-2: advancing a unified representation for a variety of vision tasks," IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR 2024), Seattle, United States, pp. 4818-4829, June 2024.

[6] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J.M. Alvarez, and P. Luo, "SegFormer: simple and efficient design for semantic segmentation with transformers," Neural Information Processing Systems (NeurIPS 2021), vol. 34, pp. 12077-12090, Dec. 2021.

[7] A. Kodaira, C. Xu, T. Hazama, T. Yoshimoto, K. Ohno, S. Mitsuhori, S. Sugano, H. Cho, Z. Liu, M. Tomizuka, and K. Keutzer, "StreamDiffusion: a pipeline-level solution for real-time interactive generation," IEEE/CVF Int. Conf. Comput. Vision (ICCV 2025), Honolulu, United States, pp. 12371-12380, Oct. 2025.

[8] F. Reda, J. Kontkanen, E. Tabellion, D. Sun, C. Pantofaru, and B. Curless, "FILM: frame interpolation for large motion," European Conf. Computer Vision (ECCV 2022), Tel Aviv, Israel, pp. 1-19, Oct. 2022.

[9] S. Hua, Y. Sun, K. Ma, D. Niyato, and M. A. Imran, "A mathematical framework of semantic communication based on category theory," arXiv:2504.11334, April 2025.

[10] 細沼恵里，山崎託，三好匠，田谷昭仁，西山勇毅，瀬崎薫，"生成型画像伝送手法におけるエッジ抽出を用いた構図情報の再現性向上，" 2024 信学ソ大, B-6-48, p. 48, Sept. 2024.

[11] T. Zhang, V. Kishore, F. Wu, K.Q. Weinberger, and Y. Artzi, "BERTScore: Evaluating text generation with bert," 2020 Int. Conf. Learning Representations (ICLR 2020), pp.1-43, April 2020.

## 〈発　表　資　料〉

| 題　　名 | 掲載誌・学会名等 | 発表年月 |
|---|---|---|
| 意味情報に基づく生成型画像伝送におけるシーングラフの活用について | 電子情報通信学会技術研究報告 | 2025 年 10 月 |
| ［依頼講演］マルチモーダルな生成型セマンティック通信における意味的類似度に基づく復元画像の評価指標 | 電子情報通信学会技術研究報告 | 2025 年 10 月 |
| ［招待講演］マルチモーダルな生成型セマンティック通信における構図情報抽出手法の比較と再現性評価 | 電気学会通信研究会 | 2025 年 12 月 |
| Applying Generative Image Transmission Method Based on Semantic Information to Video Transmission | 2026 RISP International Workshop on Nonlinear Circuits, Communications and Signal Processing (NCSP 2026) | 2026 年 2 月 |
| 生成型画像伝送におけるシーングラフを用いた意味情報の選択 | 2026 年電子情報通信学会総合大会 | 2026 年 3 月 |
| 意味情報に基づく生成型画像伝送システムの初期実装 | 2026 年電子情報通信学会総合大会 | 2026 年 3 月 |
| ［ポスター講演］体感品質に基づく生成型映像伝送アーキテクチャの初期検討 | サイバーライフラインに関する分野横断型研究会 | 2026 年 6 月 |
| 意味情報に基づく生成型動画伝送に関する初期検討 | 電子情報通信学会技術研究報告 | 2026 年 6 月 |
| ［依頼講演］意味情報に基づく生成型画像伝送の概要と動画伝送への応用 | 電子情報通信学会技術研究報告 | 2026 年 7 月 |