G11

t-Room における赤外線センサを用いた 動画像オブジェクト抽出

Moving Image Object Extraction with Infrared Sensor for "t-Room"

上埜 敏司†　片桐 滋†　大崎 美穂†
Satoshi Ueno　Shigeru Katagiri　Miho Ohsaki

1. はじめに

通信技術のめざましい発展と共に，電話やテレビ会議などの対話型コミュニケーションを支援するシステムだけではなく，遠隔地間の共同作業を伴う遠隔コラボレーションを支援するシステムへの関心が急速に高まっている．こうした中で，大型ディスプレイやスピーカー等で部屋を構成し，部屋全体を重ね合わせることで視聴覚メディアの対称性，いわゆる同室感を高めて遠隔コラボレーションの支援を行う t-Room が提案されている[1]．しかしこの t-Room には，ディスプレイとカメラ自身が対峙する特殊な構造を採ることに起因して，ディスプレイ上の映像に同じ撮影映像が繰り返し重ね合わさって表示される映像エコーの問題が残されている．

これまでのところ，実験用の t-Room システムでは，この映像エコー問題はカメラに偏光フィルターを付けることで回避することが試みられてきた．しかし，この手法は，液晶ディスプレイの使用に限定されるものであり，かつ偏光フィルター利用の結果として映像が暗くなる問題なども伴う．従って，この簡便ではあるが必ずしも十分ではない手法に代わる映像エコーキャンセリング法の確立を目指して，ディスプレイ上の表示映像からカメラによる撮影対象とその対象の背景とを分離抽出するための様々な手法が検討されてきた[2][3][4].

t-Room において撮影対象（映像オブジェクト）を抽出するということは，t-Room 壁面でもあるディスプレイの前部に位置する人物等の映像オブジェクトを，その背景となるディスプレイそのものの映像と分離して抽出することである．原理的に，こうした映像オブジェクト抽出は，例えば映像間の差分をとるなどの画像処理技術に拠ることもできる[2]．しかし，照明などの環境要因の変動に影響され，そうした画像処理のみでは必ずしも十分な抽出を実現するには至っていない[3][4].

t-Room において，撮影対象（抽出すべき映像オブジェクト）はディスプレイ壁面の前部に位置する．即ち，その対象と背景部との間には距離的な違いが存在する．従って，撮影対象の映像の抽出に測距技術を利用することは妥当なように思われる．こうして我々は，最近急速に普及し，開発環境も整っている赤外線センサ，Kinect センサによって得られる距離情報を用いて映像オブジェクトを抽出する手法の開発を進めてきた．本稿は，この提案手法の詳細と，画像処理技術に基づく従来型の抽出法との比較実験結果とを紹介するものである．

2. t-Room と従来の映像のオブジェクト抽出

2.1　t-Room の概要

多くのテレビ会議システムが隣接する部屋をつなぎあわせるように遠隔地間を接続するのに代え，t-Room は，部屋全体を重ね合わせるようにして遠隔地間を接続する．そのようにすることで，t-Room は，その利用者に視覚情報と聴覚情報の対称性をもたらし，その結果として同室感を高め，コラボレーションの質の向上を目指す.

t-Room は，それぞれ複数台の，大型ディスプレイとカメラ，マイク，スピーカーが組み込まれた壁面（モノリス）で側面を囲むようにして構成される．この際，通信を行う t-Room どうしのカメラ映像やディスプレイの位置を対称的に設定することで，視聴覚情報の対称性が実現でき，仮想的に部屋空間全体が重ね合わさるように接続される．図 1 に t-Room の外観の例を示す.

図 1　t-Room の外観.

2.2　t-Room の構成

t-Room では上記の構造を実現するために，各機能を提供するためのメディア機器に対応したサーバコンピュータが準備され，それぞれのサーバコンピュータどうしは LAN によって接続されている．t-Room どうしの接続は，これらのサーバどうしが LAN および WAN を経由して接続されることで実現される.

それぞれのサーバ機能，すなわち映像の撮影・送受信・表示や音声の収集・送受信・再生などは，t-Room 専用のソフトウェア上に実装されており，Web サーバに格納されている制御用 XML ファイルを読み込むことで制御される．特に映像メディアに関しては，Camera Server と Display

† 同志社大学，Doshisha University of Information Technology

Server があり，また映像データの編集機能なども持つ Proxy Server がある．図 2 に映像データにかかわる t-Room のサーバ構成を示す．

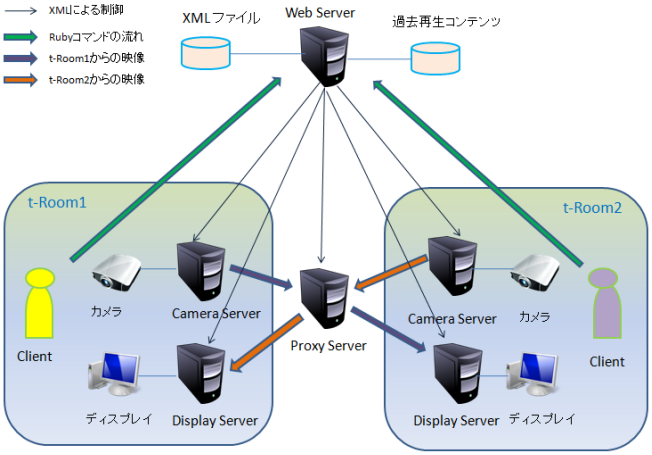

図 2　t-Room の映像データ通信に関するサーバ構成．

2.3　t-Room の映像エコーの問題

t-Room では，カメラとディスプレイが対峙する位置に配置される．このような構成のシステムではカメラとディスプレイのデータ伝送の閉ループが出来上がるため，映像エコーが発生してしまう．以下の図 3 に映像エコーが生じるメカニズムを図解する．

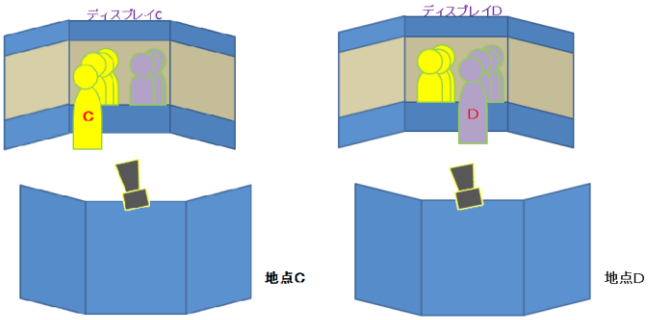

図 3　t-Room の映像エコーメカニズム

図 3 では地点 C と地点 D の t-Room が接続されている場合に，地点 C でディスプレイ C と対応したカメラが対峙しており，地点 D でディスプレイ D と対応したカメラが対峙している場合を考えている．地点 C のカメラはディスプレイ C の前に立つ被写体の人物を撮影しており，その映像は地点 D のディスプレイ D に表示されている．同様に，地点 D のカメラはディスプレイ D の前に立つ人物を撮影しており，その映像は地点 C のディスプレイ C に表示される．その結果，地点 C のカメラはディスプレイ C に映っている地点 D の前の人間も撮影することになり，その地点 D の人物の映像はディスプレイ D に重ねて表示される．同様のことがディスプレイ C と地点 C の人物にも起こるため，映像信号は閉ループ上を永遠に繰り返し伝えられ，その結果として映像フィードバックによるエコーが発生する．以下の図 4 が映像のエコーが発生している状態である．

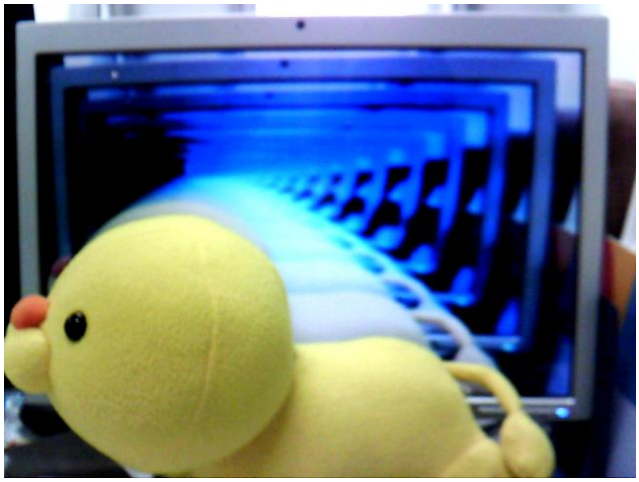

図 4　t-Room の映像エコー．

2.4　画像処理による映像オブジェクト抽出法

従来の t-Room では，問題となっている前節の映像エコーを生み出す原因であるカメラとディスプレイが対峙する t-Room 独自の構造を利用し，カメラ映像とディスプレイ映像の間で差分をとることで，映像オブジェクトを抽出してこの問題を解決していた．これは実際に映像エコーが発生する領域がディスプレイが撮影されている領域であることに注目した手法である．t-Room のディスプレイ・サーバはディスプレイに表示されている映像を知ることができる．従来の映像オブジェクト抽出法では，これを利用して取得したカメラ映像からディスプレイ映像の差分をとることにより，ディスプレイの前にある撮影対象のみを抽出して表示することができる．しかし，t-Room におけるカメラの映像は一般的にディスプレイ映像に対して，撮影する領域内にディスプレイの外側を含んでおり，カメラの角度によってディスプレイ映像に歪を含んでしまう．さらに，光の当たり方によってディスプレイに表示されている映像の色情報とカメラで撮影した色情報と一致しないという問題がある．従って，カメラ映像とディスプレイ映像を一致させ，映像間の差分から映像オブジェクトを抽出するには，毎フレーム毎にディスプレイ映像に対して，次の前処理を行う必要がある．

- **位置変換**：カメラ映像に対してディスプレイが映る有効範囲を特定し，この範囲に合わせてディスプレイ映像を変換してカメラ映像に一致させる
- **色変換**：ディスプレイに表示する色情報を変換し，カメラで撮影して得られる映像の色情報に一致させる

この 2 つの前処理を行うことで，ディスプレイ映像は映像が有効領域内に収まり，色情報が一致するため，差分によりディスプレイ映像をカメラ映像の有効範囲から消去することが可能となり，カメラ映像から映像オブジェクトが抽出される．

図 5 に従来の画像処理による映像オブジェクト抽出法の概要を図示する．

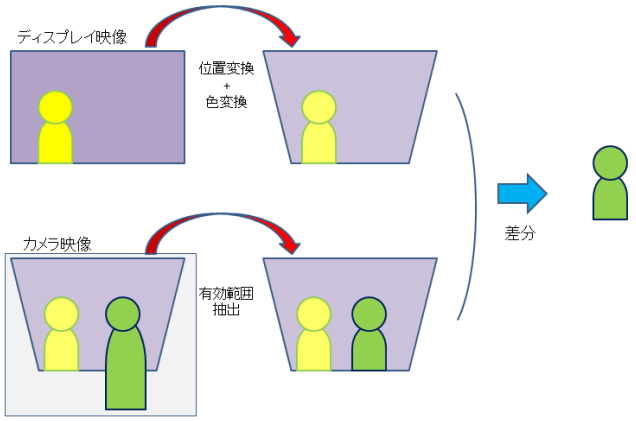

図 5　画像間の差分による映像オブジェクト抽出法.

このとき前処理部で明確にした映像間の差分から，座標と色情報の対応関係をそれぞれ位置変換テーブルと色変換テーブルとして作成しておく．これらの変換テーブルは画素値に変換情報を格納し，画像ファイルとして保存され，このテーブルを用いることで対応している座標などに変換する際に必要な計算を省略することができる．以下の図 6 に位置変換テーブルと色変換テーブルの作成例を示す.

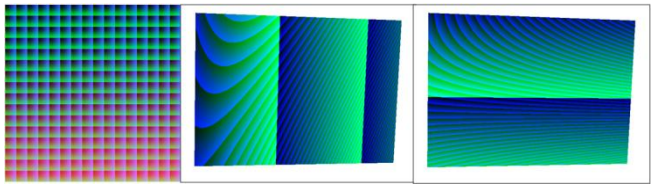

図 6　色変換テーブル（左）と位置変換テーブル（真中，右）.

しかし，この手法では映像オブジェクトの抽出精度が未だに十分ではなく，また映像オブジェクトが移動したときの映像の残像やディスプレイ映像と重なった場合に映像エコーが発生してしまう．このことから，この問題を解決するためにより一層の工夫，もしくは新たな映像オブジェクト抽出法が必要であると考えられる.

3. Kinect センサ

3.1　概要

Kinect センサは Microsoft 社の提唱した新しいインタフェースの概念である NUI（Natural User Interface）を実現させるためのツールとして開発された．もともと家庭用ゲーム機の Xbox360 用ゲームコントローラとして発売され，プレイヤーの身体の動きや声を感知して，その情報をリアルタイムにゲームのプログラムと連携させることができるというものであった．Kinect の特徴として，搭載されている赤外線センサによってリアルタイムに使用者の距離情報を取得して処理をすることが可能である点がある．これにより特別な施設や道具を使用しなくても生身の人間の姿勢や動きを検出することが可能となり，プレイヤーが何も持たずにゲームを直観的に楽しむことができるようになった．その汎用性の高い機能から，Kinect は発売後多くのユーザによって解析が進められ，ユーザの要望に応える形で PC 上で動作させるためのオープンソースドライバが開発され公開された．Windows PC で自由に開発することができる Kinect for Windows が 2012 年に発売されるなどを機に，世界中で Kinect を使用した様々なアプリケーションやサービスの研究・開発が始まった.

3.2　Kinect センサの機能

Kinect センサは，Microsoft 社の提唱した NUI を実現させるために，深度センサや映像センサ，4 つのマイクを搭載し，これらのセンサを利用して主に以下の機能を提供する.

・通常のカラー映像情報の取得
・深度情報の取得
・スケルトン情報の取得
・音声方向の取得
・音声の認識

本稿で重要な役割を担うのが赤外線センサを利用した，深度情報の取得機能である．この機能を利用することで，RGB カメラによって取得できるカラー映像に対して，距離情報を利用した画像処理等を行うことができる．以下の図 7 で Kinect の外見と各センサ類を示し，センサ類の説明を述べる.

図 7　Kinect センサの外観とセンサ類.

- ◆ **近赤外光プロジェクタ**：近赤外光パターンを広範囲にレーザー照射する
- ◆ **近赤外光カメラ**：照射された近赤外光パターンを撮影し距離を計算する
- ◆ **RGB カメラ**：通常のビデオカメラと同様にカラー映像を入力する
- ◆ **マイク**：音声を認識したり，音の発生した方向を感知する.

4. Kinect センサによる「t-Room」のオブジェクト抽出機能の開発

4.1　開発環境

本研究で提案した，動画像オブジェクト抽出システムの開発環境を以下の図 8 に示す.

OS	Windows 7 Home Premium
Programming Language	C++
Integrated Development Environment	Microsoft Visual C++ 2010
Plugins	OpenNI 1.5.4
	NITE 1.5.2
	OpenCV 2.3

図 8　開発環境.

本研究では，Kinect センサに搭載されているセンサ群を利用するための API ライブラリとして，Prime Sense 社が中心となって開発・提供している OpenNI を使用する．一般に，いくつか公開されている Kinect センサのライブラリの中でも OpenNI は Kinect for Windows と並び，頻繁に利用されている 2 つの代表的なライブラリの 1 つである．Windows OS に加え，Mac OS，Linux OS，Android OS といった様々なプラットフォーム上で動作する特徴がある．また，同じくクロスプラットフォームな画像処理ライブラリである OpenCV にも対応しており，商用利用も可能なライセンスとなっている．t-Room に導入する場合には，その移植性の高さからも，OpenNI が開発環境に適切であると考えられる．

4.2　手法の概要

Kinect センサを用いた動画像オブジェクト抽出とは，t-Room で問題となっている映像エコーを発生させている RGB カメラのカラー映像と赤外線センサによって得られた距離情報から作成した奥行き画像を合わせて，動画像オブジェクトの領域を決定し，マスク処理によってカラー画像から映像オブジェクトの領域のみを抽出する手法である．重要となるのは，赤外線センサから生成する距離情報のマスク画像と RGB カメラの情報を一致させる点である．Kinect センサは前章で図示した外観からも分かるように，距離情報を出力する近赤外光カメラとカラー画像を出力する RGB カメラの位置が横にずれている．従って赤外線センサから出力された距離情報は，そのまま奥行きマスク画像に変換しただけでは RGB カメラの画像とは一致せず，ずれが生じてしまう．また，t-Room における映像は全身を撮影したものとなるため，Kinect センサに搭載されている RGB カメラ解像度は性能的に不十分である．このことから，このシステムでは外部のカメラ映像と距離情報を連携させることを想定し．位置のずれたカメラ画像と距離情報を視覚化したマスク画像を一致させる機能が必要となる．ディスプレイまでの距離を閾値とし，このずれ補正機能を利用して生成した奥行きマスク画像をカメラ画像と合わせることで，オブジェクト抽出画像を生成する．

4.3　奥行きマスク画像とカラー画像のずれ補正

前説で示した通り，本手法では Kinect センサによって生成した奥行きマスク画像とカラー画像を重ねるように，2 つの映像間のずれを補正する処理が必要である．この補正処理は Kinect センサに搭載されている RGB カメラのカラー画像と近赤外光カメラから取得した奥行き画像の間では，通常 OpenNI によって提供されている視点一致関数が利用されている．しかし，この視点一致関数は Kinect センサに搭載されている RGB カメラと近赤外光カメラの間でのずれ補正に最適化され，ハードコーディングされている．そのため外部の RGB カメラと近赤外光カメラとの間におけるずれ補正を想定していない．このことから，本研究では奥行きマスク画像とカラー画像で，映像中の対応点を 4 点認識し，対応点を元に以下の 1 式で変換される射影変換によってカメラ間のずれ補正を行う．

$$\begin{cases} X = \dfrac{a_1 x + b_1 y + c_1}{a_0 x + b_0 y + c_0} \\ Y = \dfrac{a_2 x + b_2 y + c_2}{a_0 x + b_0 y + c_0} \end{cases} \tag{1}$$

ここで注意を要することは，対応点認識を行う際に奥行き画像は近赤外光カメラから入力された距離情報をもとに 0 から 255 の範囲に正規化し，モノクロ画像として視覚化したものであるということである．この奥行き画像は，通常の RGB カメラが撮影するカラー画像とは全く別の情報である距離情報を撮影して作成されている．従って，RGB などの色の情報によってのみ認識される線やマーカーなどを対象とした画像認識では，奥行き画像上では認識することはできない．この問題を解決するために，本研究では，通常のカラー画像と奥行き画像の両方での認識を考慮し，対応点の認識に色が均一である板状のものに円形の穴をあけて，カラー画像の色情報はもちろん，奥行き画像の距離情報においても同様の輪郭を画像上に描画できるものを認識の対象とした．この対象物体を撮影した 2 つの画像に対して Hough 変換を適用することで，円の中心点を 4 点検出する．以下の図 9 に実際にカラー画像と奥行き画像のそれぞれで，4 つの対応点として認識している状態を示す．この図では，各映像ごとに円形輪郭とその中心座標を対応点として認識し，描画している．

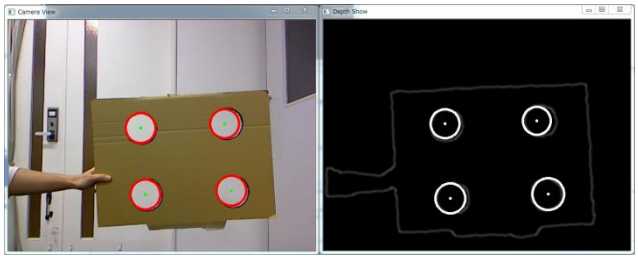

図 9　対応点の認識結果.

この認識した対応点座標を使用し，1 式を用いて射影変換行列を生成する．この行列によって奥行き画像全体を変換することで，2 つの画像間のずれを補正することができる．実際にこの行列を利用し，奥行き画像とカラー画像のずれを補正した結果が下の図 10 である．なお補正した奥行きマスク画像（左下）とカラー画像（左上）を利用して実際にオブジェクト抽出を行った映像が右下の映像である．

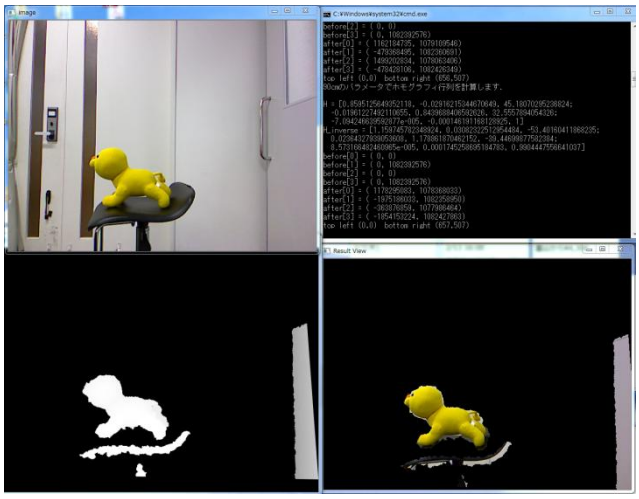

図 10　動画像オブジェクト抽出結果.

5. オブジェクト抽出の精度評価実験

5.1　目的と実験条件

本研究で開発した Kinect センサを用いた動画像オブジェクト抽出法と従来の t-Room で利用されていた映像差分によるオブジェクト抽出法との比較実験を行った．性能の比較は特に映像オブジェクトの抽出精度に焦点を当てており，以下の条件をもとに精度を求める．

- 本研究で提案した手法と従来の映像差分による手法の２つを比較する
- t-Room での使用を想定し，オブジェクトの後ろにディスプレイを設置する
- 後ろに設置したディスプレイを用いて，従来手法に必要な前処理を行う
- オブジェクト抽出精度を視認しやすくするためにディスプレイに赤色のカラー画像を表示する
- RGB カメラから静止画を撮影し正解画像を作る
- 正解画像との差分によって抽出誤差をピクセルごとに判定する
- 映像差分を利用した手法の色変換誤差による抽出ミスを軽減するために　色の差分閾値を設定する

なお，本研究では 2 つの手法を比較する際に，ピクセルごとのオブジェクト抽出誤差を次の 2 つの指標に使用し，それぞれの指標での評価結果を使用し，提案手法と従来手法の特徴を考察し，本研究で提案した手法の有用性を検討する．

- **False Addition**：正解画像中における背景ピクセルが，オブジェクト抽出画像においてオブジェクトとして認識され，抽出されている割合.
- **False Deletion**：正解画像中におけるオブジェクトピクセルが，オブジェクト抽出画像において背景と認識され，抽出されていない割合

5.2　結果

実験では，はじめに RGB カメラで，撮影対象となるオブジェクトの後ろに設置した赤と青のカラー画像を表示するディスプレイを撮影することで，オブジェクトの有効領域を認識した．この有効領域は，従来手法における位置変換を行ったディスプレイ映像と，カメラ映像の差分を取る際に利用し，前節で定義した False Addition と False Deletion を求める際にも必要となる．以下の図 11 に赤のカラー画像を撮影した結果と，認識した有効領域を出力した結果を示す.

図 11　赤色カラー画像の表示状態と有効領域.

次に撮影対象となるオブジェクトをディスプレイ前に設置しオブジェクト抽出の正解画像を作成する．以下の図 12 に実際にオブジェクトを設置した状態のカラー画像とオブジェクト抽出の正解画像を図示する.

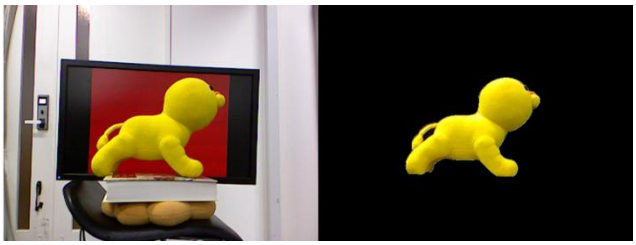

図 12　カラー画像と正解画像.

上の図に対して，実際に本研究で提案した手法を用いて，映像オブジェクトを抽出した結果が下の図 13 である.

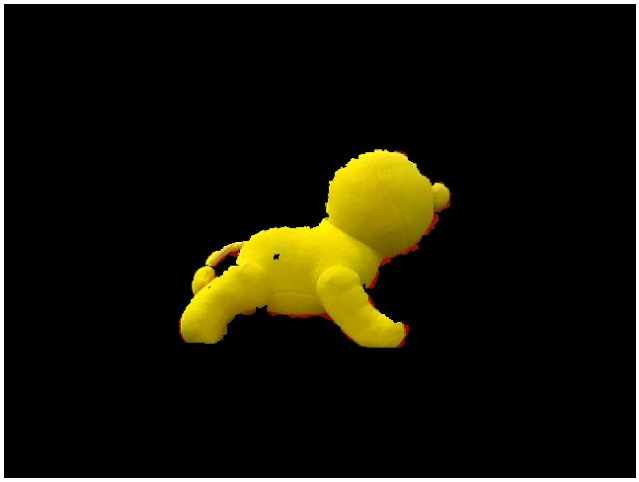

図 13　提案手法のオブジェクト抽出結果.

この結果を見ても分かるように本手法で開発したシステム抽出結果では，抽出したオブジェクト領域に抽出対象の他に，背景である赤色の領域が含まれている．これは Kinect センサの近赤外光カメラから作成した奥行きマスク画像とカラー画像でのずれが完全には補正しきれていないことが主な原因と考えられる．また，赤外光の反射を撮影することで得る距離情報では，撮影するオブジェクトの形状によっては輪郭部付近で反射の角度が大きくなり不安定に変動していることが考えられる．

次に比較する，従来型の映像差分によるオブジェクト抽出法を適用した結果が下の図 14 である．

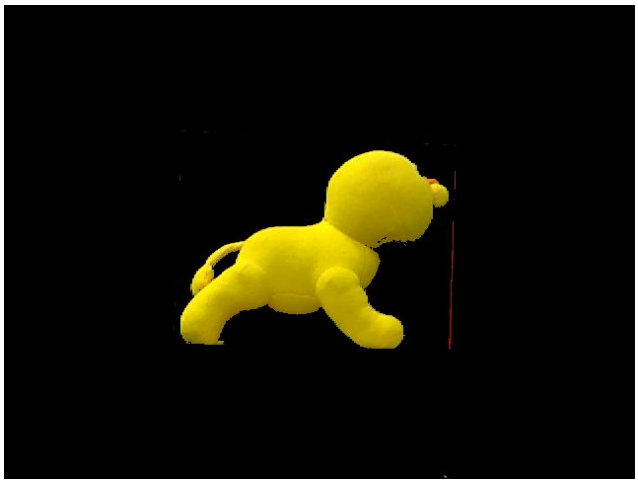

図 14　映像差分による手法のオブジェクト抽出結果.

上図の結果を見てみると，抽出するオブジェクトの色情報はその場の明るさによって大きく変わっていることが分かる．そのためディスプレイに出力する画像とカメラで撮影する画像の色情報には照明などの環境によって大きな変化が起こる．これが従来法における色変換の誤差につながり，実際にオブジェクトの抽出ミスが起こっている．以下の図 15 に，前節で定義した評価指標を用いて提案したオブジェクト抽出法と従来のオブジェクト抽出法に対する評価した結果を示す.

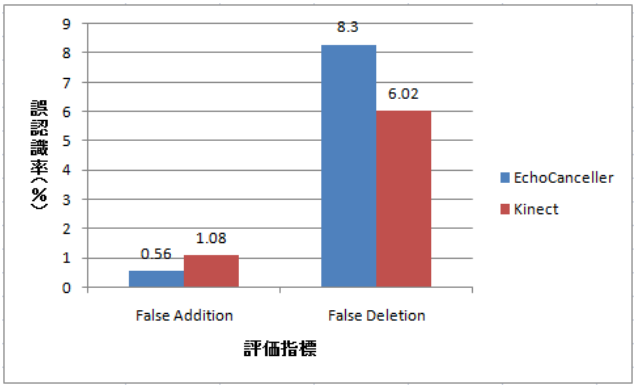

図 15　オブジェクト抽出評価結果の比較.

上図では，青色の EchoCanceller という棒グラフが t-Room で使用されていた従来のオブジェクト抽出法の評価を表しており，赤色の Kinect の棒グラフが本研究で提案した手法の評価結果を表している.

5.3　考察

前節の評価結果から，本研究で定義した 2 つの評価指標の内，提案したオブジェクト抽出法における False Deletion の評価では，従来の画像処理を利用したオブジェクト抽出法を用いた結果よりもオブジェクト抽出の精度が優れており，False Addition の評価では，精度がわずかに劣っているということが分かる．False Addition は定義の通り，オブジェクト抽出の有効領域によって値が変動するため，有効領域が小さい領域となったなら，False Addition の値が大きくなる場合が考えられる．しかし，t-Room での実行環境においては，カメラ映像中のディスプレイ領域はできるだけ大きく設定することが好まれることから，有効領域が本実験と同等か，さらに大きくなると考えられる．よって提案した手法は従来法と比較すると同等のオブジェクト抽出精度を持っているということが言える．さらに赤外線を使用することで，照明などの影響を受けずにオブジェクト抽出を行うことができる．また実行する際の前処理も手順を減らすことができた．しかし，オブジェクト抽出結果の輪郭部付近の抽出が失敗していることから，カラー画像と奥行き画像のずれ補正が不十分であることが分かる．この画像間のずれ補正精度を，より向上させることでさらにオブジェクト抽出の精度を向上させることが可能だと考えられる.

6.　おわりに

6.1　まとめ

本研究では，t-Room の環境で起こる映像フィードバックによる映像エコーの解決法に着目し，Kinect センサを用いた動画像オブジェクト抽出法を提案し，実装，及び従来型の映像差分による手法との比較実験を行うことで手法の評価を行った．実際に開発したシステムを稼働させた結果，赤外線センサを利用した動画像オブジェクト抽出により，従来法で問題となっていた実行の前処理による作業の手間，外的環境による影響といった問題が改善できた．しかし，オブジェクト抽出の精度に関しては，まだ従来手法に対する優位性は十分に示すことができず，依然として改善の必要があることもわかった.

6.2　今後の展開

Kinect センサを用いた動画像オブジェクト抽出には，搭載されている近赤外光カメラと RGB カメラのずれを補正する処理が必須である．しかし，本論文でのずれ補正手法では補正の精度が不十分であるため，より一層の工夫が必要である．また，近赤外光カメラの入力から生成した奥行き画像は物体の表面の角度によって輪郭付近で距離情報の取得が不安定になる．従って，オブジェクトの抽出映像にノイズが入ってしまうため，こちらも解決するための工夫が必要である.

7.　謝辞

本研究の遂行と論文作成にあたり，終始熱心な御指導，御鞭撻を下さりました，同志社大学理工学部情報システムデザイン学科共創情報学研究室の片桐滋教授をはじめとする t-Room の研究グループの皆様に改めて感謝を申し上げます.

参考文献

[1] Keiji Hirata, Yasunori Harada, Toshihiro Takada, Shigemi Aoyagi, Yoshinari Shirai, Naomi Yamashita, and Junji Yamato; “The t-Room – Toward the Future Phone,” NTT Technical Review, Vol. 4, No. 12, pp. 26-33 (2006 12).

[2] 小寺晋平，原田康徳，平田圭二，片桐滋，大崎美穂; “映像フィードバックに伴うエコーのキャンセリング法に関する実験的評価,” 電子情報通信学会 信学技法 PRMU2009-133, pp. 291-296 (2009 11).

[3] 中村譲，小寺晋平，片桐滋，大崎美穂; “遠隔協働支援システム t-Room における映像エコーキャンセリング法の実装と実験的評価”，平成 22 年度 情報処理学会関西支部支部大会，E-13 (2010 9).

[4] 中村譲，片桐滋，大崎美穂; “遠隔協働支援システム「t-Room」における映像オブジェクト抽出法の改良”，電子情報通信学会 信学技法 ITS2011-56, IE2011-132, pp. 325-330 (2012 2).