クリッカブルなオブジェクトの撮影の一方式とその評価

菅原俊治[†] 栗原聡[†††] 青柳滋己[†]
佐藤孝治[††] 高田敏弘[†]

動画内の対象や領域に URL を対応させ，その情報を映像と同時に配信し，再生時に対象物をクリックすることで，動画から他の URL の情報を取得できるメディアがいくつか提案されている．しかし，動画の一部に選択可能（クリック可能）なエリアを指定することは容易ではない．リンクを貼り付ける対象はフレームごとに位置が変わり，フレームごとに領域を指定する作業が必要となる．本論文では，実体をデジタルカメラもしくはデジタルビデオカメラで撮影するときに，クリック可能としたい対象が写っている領域とそのリンク先 URL や撮影時刻などの関連情報も同時に取得し記録すること，およびその一実現方式として赤外線の波長差分を使う方式について提案する．さらに，その実験システムを作成したので評価結果を述べる．この実験結果から，蓄積型の映像配信だけでなく，実時間での動画像のライブ中継も可能であることが分かった．また，複数対象物のモーションキャプチャにも利用可能と考えられる．

A Method for Capturing Clickable Objects on Video Images and Its Evaluation

Toshiharu Sugawara,[†] Satoshi Kurihara,[†††] Shigemi Aoyagi,[†]
Koji Sato[††] and Toshihiro Takada[†]

This paper discusses a model and a method for capturing image maps, animations and motion pictures with clickable objects. Recently, a number of methods/systems that make objects in moving pictures clickable are proposed. Such movies and pictures are displayed in the Internet environment or broadcasted via digital TV media and their clickable objects are linked with appropriate URLs (e.g., for shopping and for getting more detail information). It is, however, a hard task to specify the clickable areas of objects in video because they move around over frames. To cope with this problem, this paper proposes a method for taking/capturing a (motion) picture with associated information of and the data of the clickable areas of some objects that the creator wants to be clickable. The proposed method uses the difference of wavelengths of the infrared-band spectrum, with collateral information associated with objects. Then we also show the experimental results using our prototype system. These results suggest that our method can be realized as a real-time system. Our proposed method is also applicable to motion-capture systems.

1. はじめに

静止画を Web のページで表示し，その部分領域を選択可能（クリック可能）とし，あらかじめ指定された URL の情報を提示できる．動画についても，Web を含めたインターネットや，インターネットと結びついたデジタル TV 放送による動画の配信・再生において，動画内の特定の領域もしくは特定の対象をクリック可能とし，実際にそれをクリックすると URL などで指定された関連情報を提示したり，プログラムの起動などのイベントを発生させたりすることが可能となってきた[1),3),4)]．

しかし，実際に，動画の一部にクリック可能なエリアを指定することは容易ではない．リンクを貼り付ける対象はフレームごとに位置が変わり，各フレームで領域を指定する作業が必要になる．我々はこのような問題に対処するために，実体をデジタルビデオカメラで撮影するときに，クリック可能としたい対象が写っているイメージ上の領域と，そのリンク先 URL や撮影時刻などの関連する情報も同時に取得/記録/再生す

† NTT コミュニケーション科学基礎研究所
NTT Communication Science Laboratories
†† NTT サイバースペース研究所
NTT Cyber Space Laboratories
††† 大阪大学
Osaka University

るモデル，および，その一実現方式について提案している[12]．本方式で撮影した映像をブラウザなどで提示するときに，クリック可能な対象の範囲やリンク先URLなども，撮影時に埋め込まれた関連情報から得ることができる．この考え方は静止画の場合にも有効である．現状では，写真などを専用エディタで，写真ごとにクリッカブルとしたい対象物の領域を指定する．枚数が多くなればその作業の負荷も増す．本方式を静止画に適用すると，瞬間的な撮影によりイメージマップが構成できる．

同様なことを実現する研究として，現実の世界にテキストや映像などのデジタルデータを埋め込む拡張現実感（augmented reality，AR）がある[2),6)]．ARでは，実世界の映像を計算機を通して付随する情報や仮想的な物体像を重ね合わせて映し出し，情報の添付やエンタテインメントを実現する．通常は，情報が添付されるところに位置と対象物の同定のためのタグや装置が置かれており，その情報に基づいて付加情報の出現位置を決めている．一方，我々の方式で作られる映像は人間の目に見えたとおりであり，タグや装置など特別なものは写らないか，人が見る限り気づかない程度の十分小さいことを基本方針とする．さらに，対象物の在りかではなく，撮影した映像内での領域を同定することに主眼を置く．もちろん，生成された映像と関連情報を使って，テキストや映像をはめ込むことは可能であるが，ここでは言及しない．むしろ映像は見た目では自然なものであり，その中のクリッカブルとしたい対象の領域を純粋に同定することを目的とする．このように作られた映像をインターネットで受信しブラウザやプレイヤで再生するか，デジタルTVとしてケーブルや衛星で配信し視聴しているときに，そこに映っている対象をクリックすることで，何らかのイベントを発生させることが目標である．一般のARの研究のように，対象物ごとの多種類のIDをつけて区別するのではなく，撮影のシーン・カットあるいは場所ごとに映像内の対象物とその領域を区別することに主眼を置いている．

しかし以前，我々が報告したシステムは，原理確認の実装であり，十分な評価を行っていなかった．また，いくつかの問題点があると考えられていた．その1つに，本方式では4つのDV形式の映像を利用したため，実際のネットワーク配信や実時間性を考えると，その実現が難しいと考えられていたことがある．そこで，我々はDVのソフトウェアデコーダを開発し，それを利用して実際に提案方式の実装を行い，評価実験を試みた．その結果，毎秒約30フレーム以上の撮影と再生が可能であり，十分に実時間性があることが分かった．本論文の目的は，我々の方式の基本概念を説明し，実験結果を反映させたうえでのシステム構成を述べ，さらに，実時間性に関する評価実験結果について述べることにある．

本論文では，まず，本研究の目的とシステム概要を述べ，その後に要求条件を満たす一方式を提案する．この方式は，赤外光の波長差を利用するものである．最後に，実験により提案手法が実際に機能すること，およびシステムパフォーマンスの評価結果を述べる．

2. 実体のクリック可能性と撮影

2.1 問 題 点

本論文の一部の筆者らは，Webの空間において，つねにリンクの終点であった動画をリンクの起点にもなりうるように，Cmew（Continuous media with the Web）を提案した[15)]．Cmewでは，MPEG system streamにリンク情報を埋め込み，映像と同時に配信することで，動画から他のURLへのリンクを可能とする．またプロキシなどを利用し，リンク情報を特定のグループ・コミュニティ内で共有することも可能である．

しかし，一般に，動画の一部にクリック可能なエリアを指定することは容易ではない．リンクを貼り付けたい対象物は，フレームごとに画像内での位置が変わり，実際には各フレームごとに領域を指定する作業が必要になる．この問題に対処するために，たとえばCmewでは，簡易な画像認識アルゴリズムを導入し，あるフレームで対象物の領域を指定すると，それ以降は対象物の移動に合わせて領域を追随させるよう試みている[16)]．しかし，現状の画像認識技術は精度の面で十分に強力でなく正確な物体の追跡はできないか，逆に精度の高い認識のためにはかなりの処理時間を必要とする．このためにVHM[11)]では物体の追跡は行わず，数フレームごとに人間が領域を指定し，線形補間で近似する方式もとられている．しかし，対象物はいつも線形的に動くわけではないので，やはり人による注意深い編集が必要となる．

たとえば**図1**の例を考える．(a)のように大きさや向きが変わると簡単な画像認識では追随が難しい．照明や影などの影響も受ける．より精密な認識技術もあるが，その分コストの高い処理が要求される．(b)のような場合には，オブジェクトが領域内を線形に移動していないので線形補間がうまく適用できない．また，(c)の場合には，部分だけが映っているフレームがあり，単純な認識技術や線形補間が使えない例である．

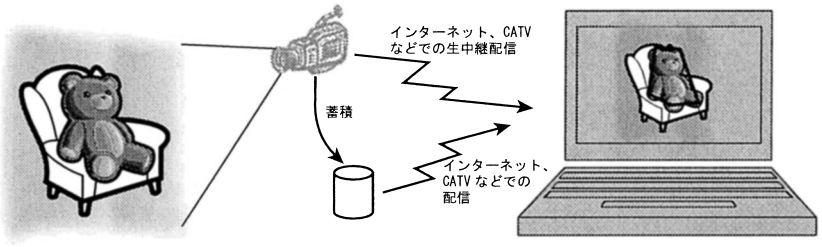

図 2 実体の撮影から再生までの流れ
Fig. 2 From video recording to playback.

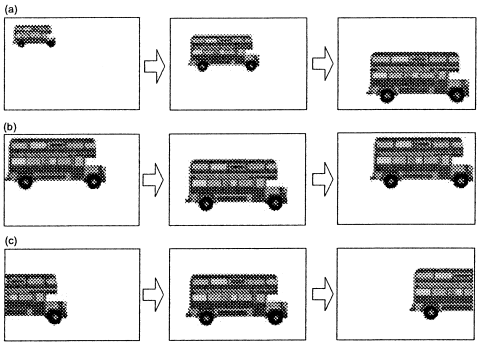

図 1 オブジェクトの撮影領域内の移動
Fig. 1 Moving objects in video recording areas.

2.2 対象物と映像のクリック

本研究の動機は，上記のような作業を軽減するために，画像内のクリック可能な領域をオーサリングツールで指定するのではなく，撮影のときに同時に決定することにある．さらに，可能かつ必要であれば，その領域に関連づける URL も同様に獲得する．ハイパーリンクが定義されている画像内のクリッカブルな対象は，仮想的にデジタル映像内の座標で指定された領域であり，そこに写った対象物を認識したり，実体と結びついているわけではない．本来は，映像内の対象ではなく実世界の実体そのものに意味があるので，実体とリンク先の情報が関連している．しかし，実世界とインターネット上のサイバー空間を直接結びつけることは現状ではできないので，写真や絵に描かれた対象物をその実体のシンボルと考え，対応させている．つまり，ユーザが映像内の対象物をクリックするのは，その写った対象物と実物と同一視し，その実物に関連するリンクをたどっている．同様に，その写真や絵の中の対象物もコンピュータが認識しているわけではなく，人間が個々の映像ごとにその対象を理解し，映像内の領域を指定して，関連する URL などのリンクを指定する作業を行っている．したがって，撮影という実体をデジタル化する瞬間に，可能な限りの関係性を持たせたい，というのが本研究の基本的な立場である．

2.3 目的と要求条件

本研究でめざすシステムは以下のとおりである（**図 2** 参照）．ある対象物を撮影するときに，その撮影された画像内での対象物の位置（輪郭のだいたいの座標）と関連情報（URL，時刻，場所など）を同時に抽出し，直接再生するか，ネットワークなどで配信の後に再生を行う．一時的に蓄積することも可能である．なお，映像情報への URL やクリッカブルな領域などの付加情報の埋込み方式は本論文の範囲外である．たとえば，文献 1)，4)，15) などが提案されており，適切なものを選択する．

このような撮影方式を実現するにあたり，以下の条件を実現する必要があると考える．第 1 に，対象物の領域が，撮影された映像にはめ込まれなければならない．つまり，映像内の対象の位置やアングルによらず，映像内の座標として取り出せる必要がある．電波のように指向性はあっても映像内の位置として取り出せないものは利用できない．さらに，カメラと対象物の距離がある程度離れていても撮影できなくてはならない．本研究では 1 m から 6 m 程度の対象物に絞った．一方で，遠方の対象物や小さな対象物をクリックするのは現実的ではないことも指摘したい．

第 2 に，実体の画像内の位置を同定するような特殊な機器や印などは，撮影した映像には不要な情報であり，見えないか事実上見えない程度の小さなものとする．たとえば，バーコードや特殊な信号を発する不自然な装置は映像上は障害物であり，写らないことが望まれる．撮影されたものは自然な画像で，これにクリッカブルな領域や関連情報を付加するのが基本的方針である．

第 3 に，ライブ的なストリーム情報の発信から，蓄積タイプの静止画・動画にも対応可能とする．特にライブ配信は文献 12) でも提案したが，技術的に不可能と考えられていた．本論文では，この問題に対する評価を行っている．

第 4 は迅速な撮影に対応することである．特に，瞬

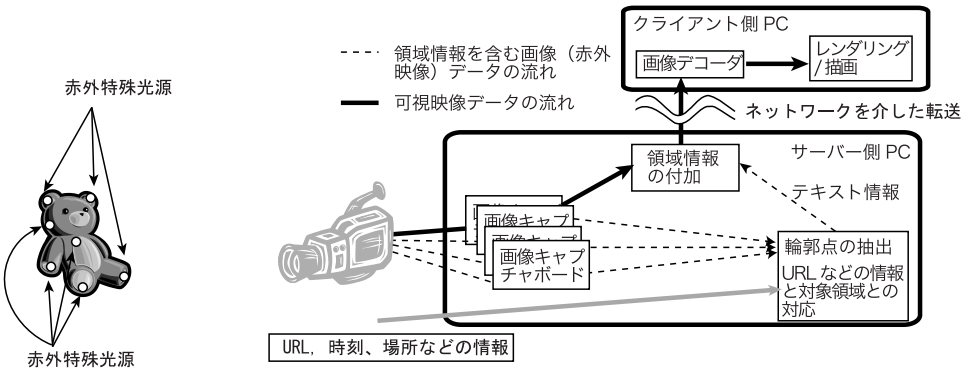

図 3 提案手法およびシステム概要図
Fig. 3 Proposed capturing method and system overview.

間的なタイミングを狙った撮影にも対応することが必要である．これは，映像作品を撮影するうえで，瞬間的なシャッタチャンスをとらえることが重要と考えるからである．

第 5 に，複数の対象物を同時に扱えることである．実際に，もし 1 つの対象物のみであれば，領域を認識せずに動画像全体をクリックする方式も考えられる．一方で，動画では多量の対象物をクリック可能とすることも考え難い．

3. 撮影方式の一提案

3.1 波長差分を利用した一実現方法

基本的なアイデアとシステムについて説明する．本方式は，**図 3** に示すように，対象物にいくつかの特殊赤外光源を埋め込み，一方，特殊カメラで可視画像に加えて光源の点を同定する映像を同時に取得する．この際，さまざまな角度からの撮影でも対象物の輪郭を同定できるように，対象物の周囲だけでなく，前や背後にもいくつかの光源をつける．これらを特殊カメラで撮影し，可視画像のほかに，特殊光源を撮影した画像を同時に収集し，システムのサーバに送る．

これらの情報を収集したシステムのサーバ側では，特殊光源の画像から，光源の画像内の位置を抽出し，特殊光源のタイプごとに分ける．この際に，同じ対象物は同じタイプの光源が装着されていることとし，同一タイプの光源を含むような領域を同定する．同時に赤外光通信，無線（無線 LAN，bluetooth），GPS などの他の方式を利用し，カメラかサーバ側で URL，時刻，位置などの関連情報を取得する．同定した領域情報と URL などの情報を対応させ，それを可視画像にあわせて蓄積するか，クライアント側に実時間配信する．このときに，領域情報などは画像に埋め込まれるか，別情報としてファイルに保存することが考えられるが，本論文ではこれらの方式については言及しない．ネットワークなどを介して情報を受けたクライアント側では，再生時にこれらの情報をあわせて提示する．

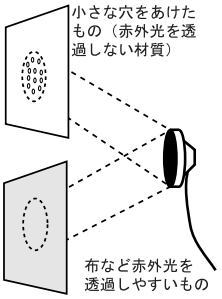

図 4 光源の設置
Fig. 4 Installation of light source.

領域指定のための光源は，撮影後も気にならない程度の小型であれば，それをクリック可能とする対象物に直接埋め込むか，**図 4** に示すように，もし 0.1–0.3 mm 程度の小さな穴があけられるなら数十個開けてその背後に置くか，布など赤外光をある程度透過する材質の背後に光源を装着する☆．クリッカブルな領域を同定するために，この光源を対象物の輪郭が浮き出るように数個装着する．

一方，画像撮影装置として，通常の CCD カメラのほかに，赤外撮影用の CCD カメラを数個用意し，光軸を合わせる．赤外撮影用のカメラには，赤外パスフィルタが装着されており，限られた範囲の波長の光のみを写す．なお，以下では 3 つの赤外撮影用 CCD カメラを使用した場合を説明する．

基本的な原理は以下のとおりである．**図 5** のように，赤外画像撮影用の CCD 側には，やや透過帯域が広い

☆ これは光源を隠すという目的もあるが，後に述べるように光を散乱させることが主目的である．

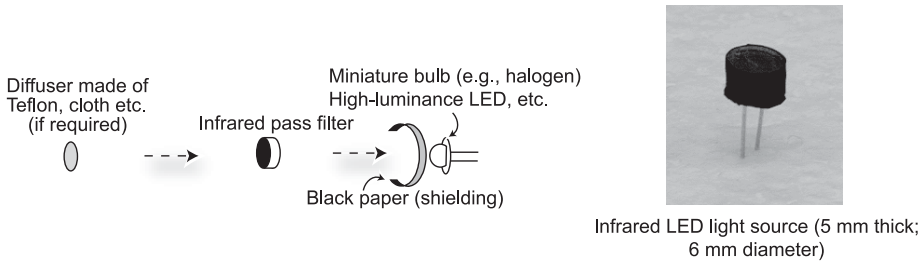

図 7 作成した LED
Fig. 7 Experimentally assembled LED.

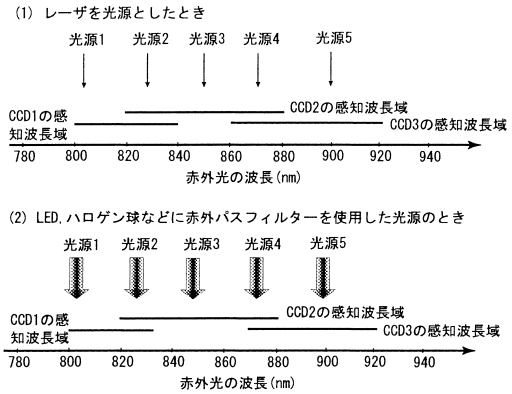

図 5 赤外カメラと光源の波長の関係
Fig. 5 Relationship between the light source wavelengths and wavelengths detected by the CCDs.

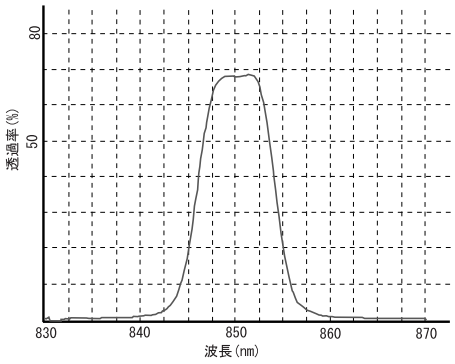

図 6 研究で用いたある赤外透過フィルタの特性（中心波長 850 nm）
Fig. 6 Example of characteristics of a narrow-bandpass filter (central wavelength: 850 nm).

赤外パスフィルタ（図 5 では，中心波長の前後 30 nm 程度）を装着する．また，光源にはクラス 1 レーザか，透過帯域の狭い赤外パスフィルタ（図 5 では，中心波長の前後 5–10 nm）を装着した LED を使用する．実際に使用したフィルタの特性を **図 6** に示す．LED の代わりに豆型ハロゲンランプなどでもよい．図 5 では，CCD に中心波長 800 nm，850 nm，900 nm のフィルタを，光源には中心波長 800 nm，825 nm，850 nm，875 nm，900 nm のものを用意し，最大 5 つまでの対象物を区別できることを示している．この場合，各光源は 1 つないし 2 つの CCD で撮影できる．3 つの CCD で撮影できた光源は，本方式のために設置した光源の光ではなく，その他の電球やその反射光または自然光などと考えられるため，領域を表す点から排除する．

取得した可視画像と赤外画像 3 種のうち，赤外画像に写っている光源の座標を各フレームごとに探し記録する．5 つの種類の光源を識別し，同じ種類の点の外郭を「ある対象物」のクリッカブル領域とする．

一方,「ある対象物」からリンクをはる URL や撮影時刻，必要ならこの対象物が何であるかなどの関連情報は，撮影した場所で，赤外線や bluetooth などで別途取り出す（3.5 節の例を参照）．なお，赤外線通信を利用する場合には領域同定に使われる赤外光と干渉を防ぐために，たとえば 940 nm や通信用の 1,000 nm 以上など 3 つの CCD で撮影できない波長領域を使うか，逆に 800–900 nm 全体を出す赤外光を通信に使用し，この部分は「他光源の排除」の操作で取り除くことなどが考えられる．この関連情報と光源の波長を照合し，対象物の関連情報として取り出す．たとえば，875 nm の映像のクリッカブル領域には http://www……にリンクを張る，というような情報を関連情報から抽出する．関連情報の詳細については，次章で述べる．

3.2 部品および装置

光　　源

実際に作成したフィルタ付き LED は，高輝度 LED に直径 5–6 mm，円板状の赤外パスフィルタ（光学実験用の高精度なもの）を取り付け周りを封印した（**図 7**）．このほかに，赤外レーザ光源に数本のファイバ接続し，各出力をクラス 1 に減光したものを光源とするものも作成した（**図 8**）．動かない対象なら，レーザポインタのようにレーザ光をあて，その反射光を撮影してもよい．たとえば歴史的展示物，貴重な資料などはこの方式を採用すると細工をする必要がない．この LED

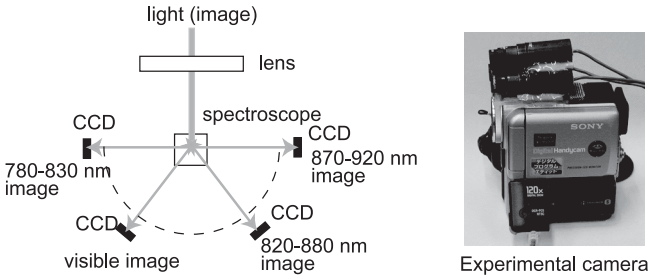

図 9 専用カメラの一例と試作したカメラ
Fig. 9 An example of four CCDs camera and experimental camera.

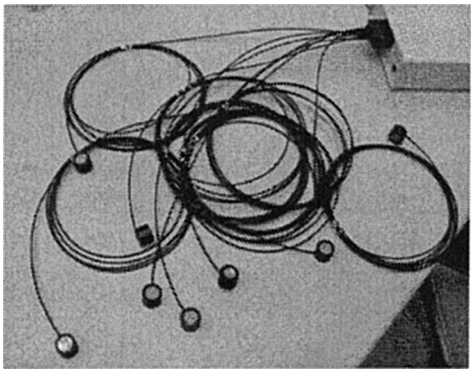

図 8 レーザ光をファイバで誘導する装置
Fig. 8 Fiber-guided laser light source.

(a)
```
<object>
  <name>string</name>           <!optional>
  <url>url string</url>         <!optional>
  <lambda>wavelength</lambda>   <!mandatory>
  <time>date-time string</time> <!optional>
  <location>location</location> <!optional>
  <script>java script</script>  <!optional>
</object>
```

(b)
```
<object>
  <id>id</id>                   <!mandatory>
  <lambda>wavelength</lambda>   <!mandatory>
  <time>date-time string</time> <!optional>
  <location>location</location> <!optional>
  <name>string</name>           <!optional>
</object>
```

図 10 関連情報の転送形式
Fig. 10 Format of collateral information related to the captured objects.

をクリック可能とする対象物に直接埋め込むか，同系色の布など赤外光をある程度透す材質の背後に装着する．また，塗料の多くはある程度赤外光を通すため，絵の具などで同系色にして目立たないようにしてもよい．クリッカブルな領域を同定するために，この光源を対象物の輪郭が形成できるように数個装着する．この際の留意点として，対象物をあらゆる角度から撮影してもだいたいの輪郭が分かるように上下左右だけでなく前や背後にも装着する．なお，ここで光を散乱させることが必須である．散乱させないと，角度によっては光が見えなくなる．布などの背後に装着する場合には必要ないが，直接光が放射される場合には，テフロン，薄い布，和紙などを光源に貼り（正確にはフィルタの外側に貼る），光を散乱させなくてはならない（図 7）．逆に，これを利用して見えないように工夫することができる．

カメラおよび映像取得装置

赤外映像作成用の CCD カメラには，光源よりやや透過する波長帯域が広いフィルタを通すか，この波長帯域のみ反射するミラーを利用し，限られた範囲の波長の光のみを写す（**図 9** 左側）．ただし，試作では複数の CCD カメラにフィルタを装着したものを組み合わせ，光軸をあわせて利用した（図 9 右側）．

4. 関連情報と抽出方式

本章では，クリッカブルとする対象物の関連情報の伝達方法と，撮影された赤外画像情報からクリッカブルな領域の抽出方式について述べる．基本的に文献 12) と同様である．

4.1 関連情報の形式

クリッカブルとする対象物の関連情報は，別途，赤外通信もしくは無線などで送信される．本例では，検出可能な赤外波長は 5 種類だけなので，ある程度位置に依存して URL などの情報を決定する必要がある．このために，たとえば赤外通信を利用した場合にはその指向性を活かし，撮影対象物の近くに置き，カメラを向けるとカメラの赤外通信用受信機が撮影と同時に関連情報を取得できるようにする．対象物の領域を検出する赤外光源に情報を載せた変調光の使用も考えられる．また，bluetooth や弱無線 LAN などの場合には，撮影エリアに入ると，情報取得可能となるようにする．

現状は，**図 10** のような情報を受信する．図 10 (a) は，

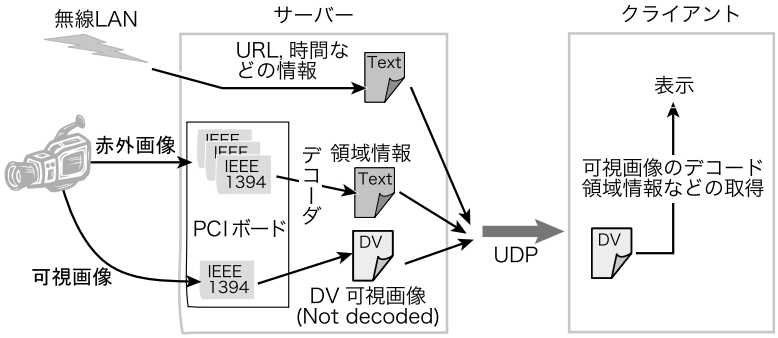

図 11 実験システム構成図
Fig. 11 System structure used in our experiments.

その環境で必要な情報をすべて取得する場合，図 10 (b) は，赤外映像の波長から ID を取り出し，後にブラウザで映像を見るときや，映像データを適当な形式に変換するときに，あらかじめ決められたデータベースなどにアクセスして URL や関連プログラムなどを取り出すことを想定している．

4.2 クリッカブルな領域の抽出と再生

赤外映像からのクリッカブルな領域を取り出すために，各フレームに以下のような単純な手法を適用する．なお，複数のカメラの光軸をあわせ，各カメラの共通部分となる画像有効領域はあらかじめ求めておく．

(1) 3 つの赤外画像において輝度の明るい点を取り出し，残りは消去する．
(2) 各波長領域で，明るい部分の塊（chunk，連結の領域）を取り出す．ある波長の chunk に対し，他の 2 波長の chunk と共通部分があるものは，他の 2 波長の chunk も含めて消去のマークをつける．
(3) 消去のマークのない chunk の形が円形と著しく異なる形であるか，ごく小さな領域か，大きな領域を占める場合には，chunk に消去のマークをつける．なお，実際の実験では DV を利用し，chunk は 3×3–25×25 ドット程度の円形に近いものとなることが分かっているので，その他の形状のものは消去マークをつければよい．さらに，この chunk と共通部分を持つ，他の波長の領域（あっても，他の 1 波長のはずである）があれば，それにも消去のマークをつける．
(4) 消去マークのない残りの円形の chunk を，写っている波長映像により，800，825，850，875，900 nm のそれぞれに分け，chunk の中心座標を求める．なお，(3) の処理の後に残った chunk は円と仮定して中心座標を求めている．
(5) 波長ごとに求めた中心点を結び，すべての中心点が境界線か領域内の内側に入る領域を求める☆．

実際には，NTSC の特性から，画像の偶数と奇数ラインを分けて処理する必要がある．また，(5) で求めた領域は，単純に外周を求めるアルゴリズムを利用しているために凸連結領域となる[7]．一般には，対象物の領域が凸とは限らないが，凸か凹かを決めるには，たとえば，光源は必ず外周にあるなどの制約情報や，対象物の形状に関する情報などが別途必要となる．

5. 実験と評価

実際に作成した実験システムと実験例を概説し，最後にシステムの評価結果について報告する．

5.1 実験システム構成

本システムを評価するために実装した実験システムの構成を **図 11** に示す．PC は，Pentium(R)4 3.2 GHz，1 GB 主メモリであり，FreeBSD5.1 上に XFree86（バージョン 4.3.0）という環境で実験プログラムを動作させた．映像入力用に，IEEE1394 のボードを 4 枚装備し，DV 形式の映像を入力している．DV のデコーダプログラムは，よく使われる libdv[5] では十分な速度が得られないので我々が開発したソフトウェア DV デコーダを利用した．

本実験システムにおいて DV ストリームとして入力される映像は 4 つある．このうち 1 つは可視画像の再生用映像であり，そのまま再生を行う．他の 3 つの DV ストリームは，赤外映像であり，これらをデコー

☆ この領域は次のように求める．ある波長の点として抽出された一番右端にある点を見つける．その点を端点とし時計の 12 時の方向へ線を引き，その線を反時計回りに回す．最初に衝突した点を次の点とする．その次の点を端点として，先の反時計回りの方法を点を発見したときの途中の角度から同様にすすめる．この操作を何度か繰り返すと最初の点に戻り，すべての外郭の点が抽出される．実際には，高速化のために，端点から見た領域を上下半面，左右半面に分け（座標の xy の値で比較し分割），探索する範囲を制限している．

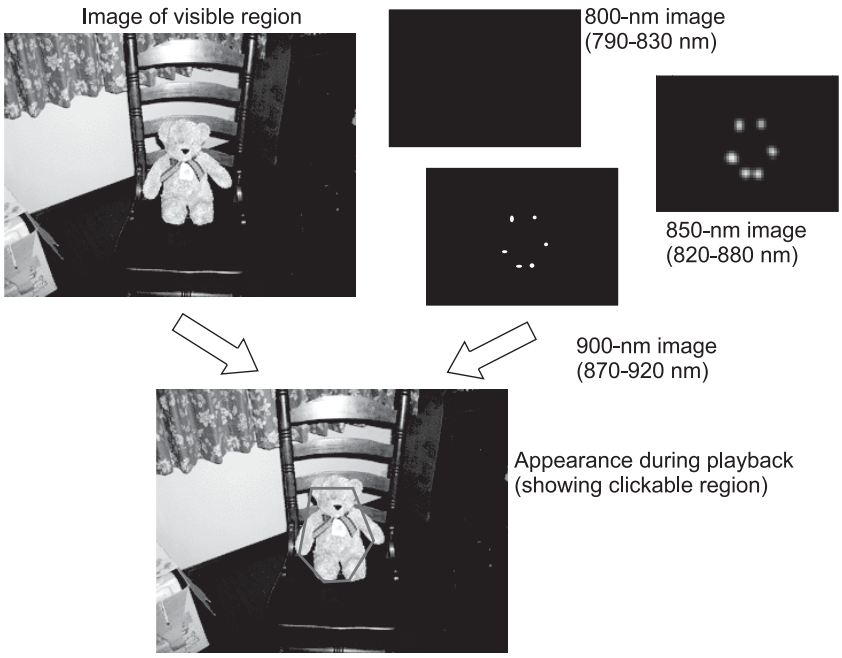

図 12 Teddy ベアの実施例
Fig. 12 Teddy Bear example.

ドし，上記のアルゴリズムを適用させ，各波長ごとの点を抽出する．この 3 映像に関しては，デコードはするが描画する必要はない．実装では各ストリームごとにスレッドを生成し，描画と点の抽出を同期させ再生した．また，再生時にクリック可能な領域を指定し，実際にクリックされると指定された URL を Web ブラウザに呼び出すことができる．この専用ブラウザは文献 12) の実験システムと同じである．

次節以降で，具体的な実験例を説明する．なお，これら各実験は，同一の部屋で行われ，照明として蛍光灯が使われている研究室の実験室で行った．この部屋に大きな窓はあるが，北向きであり，ブラインドもあるため外部からの光は少量差し込む程度である．また，窓が対象物の背後にくることを避けた．なお，本システムを文献 12) の発表の際にデモンストレーション展示を行ったことも述べておく．

5.2 実 験 例

5.2.1 基本動作確認

基本動作確認のため対象物として作成した Teddy ベアを **図 12** に示す．このぬいぐるみには，波長 875 nm の光源を 6 カ所埋め込んであるので，850 nm と 900 nm の CCD カメラにのみ写ることになる．これらを合わせて領域を抽出し，クリック可能な領域を重ね合わせてある．関連情報として受信するデータを **図 13** に示す．本例では，関連情報は無線 LAN を利用

```
<object>
   <name>teddy bear</name>
   <url>http://www.XXbear.com/</url>
   <lambda>875nm</lambda>
   <time>2003.11.6 19:21</time>
   <location>musashino lab</location>
</object>
```

図 13 Teddy ベアの関連情報
Fig. 13 Examples of collateral information for a Teddy Bear.

して受信した．本例の適用イメージは，映像に写った商品の購入を想定しており，リンク先の URL は，たとえば製造元の会社のサイトに設定する．本実験で，実際に，光源を 2 つの赤外カメラのみで写る領域を抽出し，実際にクリックすると指定した URL の情報を取り出せることを確認した．また，本例の場合，関連情報の発信元はぬいぐるみに直接埋め込むことも考えられる．

5.2.2 動く対象物での実験

ジャケットの例（**図 14**）では，移動や回転しても領域を検出できるように，900 nm の光源を前後横に 12 カ所埋め込んだ．本実験例は，動く対象物が追跡できるかを確認する目的がある．光源は，表地と裏地の間に配置，配線し，3 V のリチウム電池 2 個使用した．本例の利用イメージは，映像作品で俳優が着用し，このジャケットを着て演じた俳優をクリック可能にす

図 14 ジャケットの例
Fig. 14 Jacket example.

図 15 Teddy ベアとジャケット同時の実施例
Fig. 15 Simultaneous capturing of Teddy Bear and jacket.

ることを想定した．URL は，カットごとに異なると思われるので，カットごとに URL をデータベースなどに問い合わせることとする．ただし本実験の実装では，撮影後に撮影者が試聴，編集することを想定し，映像内の対象物をクリックしたときに URL の定義がないときに撮影者に問い合わせるようにした．実験では，実際に，このジャケットを着用し，回転や移動を行い，クリッカブルとする領域が動画にあわせて正確に指定できることを確認した．

さらに前実験のぬいぐるみとあわせて複数の対象物が検出できることも実験で確認した（**図 15**）．しかし，特に動画の場合には複数の対象物が動き重なる場合がある．本点に関しての技術的課題を 5.4 節で述べる．

5.2.3 その他の光源の排除の実験

使用した光源以外の光を排除できるかどうかを確認するために，クリスマスリースを作成した．ここで排除すべき光はクリスマスライトの光であり，この光を排除できるかを試した．本実験の **図 16** は，実際にクリスマスライトの光が 3 つの赤外映像に写り，それらが排除され，用意した LED の光のみを選択できることを示している．

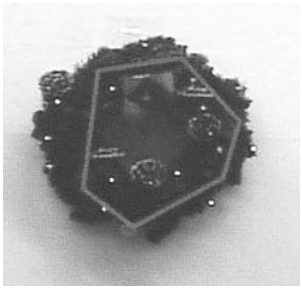

図 16 クリスマスリースの例
Fig. 16 Christmas wreath example.

表 1 実験システムの評価結果
Table 1 Experimental results of evaluation.

	点の抽出と描画	描画のみ	点の抽出のみ
FreeBSD5.1 IEEE1394（ドロップレート）	1.36	0	0
FreeBSD5.1 ディスク（処理フレーム数）	28.77	86.31	43.61
FreeBSD5.1 /VMware ディスク（処理フレーム数）	17.17	45.11	27.95

（数値の単位はすべて frame/秒）

なお，その他の光源の排除は当初の要求条件にはないが，それらを満たすための一方法として提案した本手法では，特定の波長のみを出す赤外光源を利用しており，実際に他の光源に惑わされないことを確認するために追加的に必要となった課題である．

5.3 システム評価

評価実験の結果を **表 1** に示す．以下に本実験システムの評価結果を説明する．なお，実験結果として使われている値は，同じ処理を 10 回繰り返し，その平均値とした．

1 台の PC で IEEE1394 ボード経由で映像を受信し，赤外映像から領域を抽出しそれと URL などの情報とをあわせて再生を行ったところ，見かけ上，画像の乱れはなく，自然な再生が可能であった．実際にドロップしたフレームを測定したところ，平均 1.36 フレーム/秒であった☆．さらに，実際のデコーダと再生の処理能力を測るために，可視画像と 3 つの赤外映像

☆ 1 フレームをデコードするのに必要なデータが一部でも欠けた場合をドロップフレームとした．また，ボードの順番により，ドロップの発生の偏りが若干だがある．本実験では，最大で，0.12 フレーム/秒の差が観測されたが，大きな数値ではないと考える．

をあわせた 4 つの映像をまずディスクに記録し，ディスクから読み出しながらその再生を試みた．このとき，音との同期を排除するためにオーディオ部分を OFF にして再生を行った．この実験では 28.77 フレーム/毎秒の再生速度となった．これは 30 を切るが，4 つの DV ストリームを同じディスク（ATA100）経由で取り出しているので，負荷が大きいと考えられる．なお，音の処理を OFF としたが，音の処理部分の負荷は軽い．実際，gprof で参照してもほぼ 0–1%の表示であり，上記の数値に大きく影響するとは考えられない．

評価システムの実装は，IEEE1394 を利用した場合には，各カメラごとに IEEE1394 ボードからの DV データ収集部，その DV データをデコードし必要に応じて描画する部分を別スレッドとして実装し，メインプログラムを含め全部で 9 つのスレッド構成とした．一方，ファイルから読み出す場合には，IEEE1394 ボードからのデータ取得のように同期させる必要はないので，データ収集部，DV デコーダ，描画を 1 つのスレッドとして実装し，全体として 5 スレッドの構成とした．

上記の実験では，2 章で述べた想定される利用において，サーバとクライアントが同一マシン上にある場合に相当する．実際には，図 3 の波線部分でネットワークを介する形で分割され，サーバとクライアントが別マシンとなる．また，処理時間のほとんどが DV のデコードと描画で費やされる．このため，サーバ側では，描画はないが 3 つの赤外映像の DV ストリームのデコードとネットワークからのデータ送信，クライアント側では，1 つの可視画像の DV ストリームのデータ受信，デコード，描画であり，負荷もサーバとクライアント間である程度分散される．実際に，上記のマシンで，クライアント側のみの処理を行うと，ディスクから読み取ったものは，86.31 フレーム/秒の処理能力を持ち，サーバ側のみの処理を行うと 43.61 フレーム/秒であった．またこれを IEEE1394 カードから直接読み取った場合，サーバ側の処理でも，またクライアント側の処理でもドロップフレームは発生しなかった．

クライアント側の計算機を意識して，本実験の一部をノート PC （ThinkPAD T40，Pentium(R)M，1.6 GHz，メモリ 1 GB）に VMWare（バージョン 4.0.0，メモリ 256 MB をアロケート，ホスト OS は Windows XP Professional）を載せ，その上の FreeBSD5.1（XFree86 は 4.3.0，ただし XV は利用せず）で再生のみを試みた．その結果，クライアント側に相当する再生のみでは 42.9 フレーム/秒の処理速度であった．

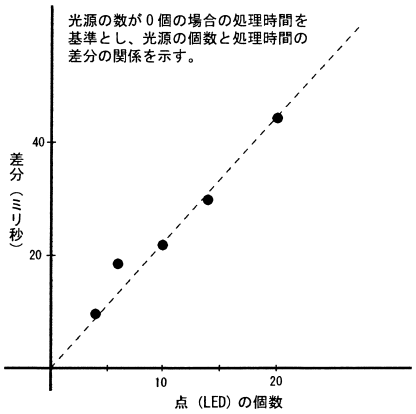

図 17 光源の数と処理速度の差分（30 フレームあたり）
Fig. 17 Number of light source and performance (every 30 frames).

最後に，本アルゴリズムは赤外映像に写った点の個数に影響を受けるので，光源の数と処理速度の差分を測った．結果のグラフを **図 17** に示す．本グラフが示すように，点の個数に対してはぼ線形に処理速度が増すことが確認できる．しかし，処理の多くは DV のデコードや表示部分が占めるため，相対的割合は大きくない．実際に，本グラフから，100 個の光源を使ったときのオーバヘッドは，1 秒間の描画フレーム数である 30 フレームにつきおよそ 222 ms と推定される．本処理はサーバ側で行うものであるので，表 1 の値と勘案すると，100 個の光源がある場合には毎秒 32.97 フレームの処理能力と考えられる．一方，実質的に光源を 100 個以上含む映像を処理することはないと考えている．これらの評価結果から，本方式は現状の技術で十分に実時間性が確保でき，ライブ中継も可能と考えられる．

処理速度が，実時間に耐えうるまで向上した工夫点を以下に述べる．第 1 にソフトウェア DV デコーダ自体の高速化がある．これは，本システムとは独立した共通ライブラリとしての工夫であり，詳細については本論文の範囲を超えるのでここでは割愛する．第 2 点は，赤外光は基本的に白黒なので，Y，Cb，Cr のうちの Cb，Cr の計算を省略している．第 3 に，通常は，DCT ブロックの復元に 3 回のデータ処理が必要だが，点の検出に鮮明な画像は不要であるので，これを 1 回に制限した．具体的にはデータの高周波の部分の復元を省略している．詳しくは，DV 標準仕様書[8] の処理ビデオセグメントの再配置アルゴリズム（arrangement algorithm of video segments）を参照されたい．さらに，上記の実験システムでは行ってい

ないが，処理速度の向上や IEEE1394 経由のドロップフレームを防ぐために，赤外映像のデコードや点の抽出を 1 フレームおきとしても大きな問題はなく，これによりパーフォーマンスは格段に向上すると予想される．

5.4 要求条件と課題

2.3 節に述べた要求条件と本提案システムの関係について概説する．まず第 1 と第 2 の要求条件である画像内の座標で位置が取り出せること，不要なタグなどの機器は見えないか十分小さい，という条件は完全でない部分もあるがある程度は満たしていると考える．また実験で使った対象物は 6 m ほど離して撮影しても輪郭が同定できることも確認した．第 3 の実時間性であるが，サーバとクライアント側に分けた処理を行うと，100 程度の光源でも十分な処理速度が得られることを実験で示した．また，光源は，対象物の輪郭をとるために使われるので，一般にはその形状や対象物が移動・回転するかなどにもよるが，実際面を考えると精密な輪郭を同定する必要はないと考えており，おおよそ 20 もあればたいがいの外郭を同定できると考える．ただし，より精密化させるために，たとえば同じ色の領域は拡張するなど，簡単な画像認識処理を追加する方法も考えられるが，本論文の技術とは独立であり，ここでは触れない．

第 4 の要求条件であるが，本方式では点滅などの変調信号を同定する手法はいっさい使っていない．このため，信号により ID を確認したり，点滅で消えているのかそれとも影に隠れて消えているのかを区別したりするなどの処理は必要としない．より具体的には，毎秒 30 フレームの映像を利用して点滅を検出し，ID を収集するには，おおよそ 30 ms × ID の長さ分の時間がかかる．また点滅で消えているのか，障害物で見えないのかを判定するのにも時間がかかる．しかし，本方式では点滅などを使用していないため，動画であっても瞬時に（1 フレームで）同定できる．関連情報については，より高速な通信手段を利用できるので，実用上も問題は発生しないと考える．

第 5 の複数対象物の同定であるが，本方式で明らかなように，5 つまでの対象物を区別できる．実際に，実験でも複数対象物が分離・同定されることも確認しているが，重なりあう対象物の選択には，より一般的な問題があることも指摘したい．どちらが手前側にある対象物かを多くの場合は同定できるが，特別な情報がない限り一般には可能ではない．たとえば，(1) 一方の輪郭の中に他の対象物の光源がある，(2) 動画の数フレーム前での情報から位置を推定する，(3) 対象物の情報，たとえば色が既知である，などの方法や情報を利用して，前後関係を同定できる場合がある．本論文の提案方式だけでは，(1) の観点から部分的情報は与えるが完全ではなく，課題として残っている．

本方式では，動かない対象物ではレーザの反射光を使うことができるが，その他の場合には小型のデバイスを対象物の背後などに配置する必要があり，適用範囲を制限する可能性はある．この意味で現状の技術では十分でない部分もあり，特に光源の小型化は必要と考えている．本方式では，点滅などの必要がなく，特殊な回路や配線も必要ない．LED と電池だけから構成され，さらなる小型化は可能と考えているが，本点も今後の課題としてあげておきたい．

5.5 実験再現のための注意点

最後に本実験システムを再現するための注意点について述べる．第 1 に，本実験では，市販の白黒 CCD カメラを利用した．そのために，900 nm のフィルタをつけた映像はそのほかのフィルタをつけたものより暗くなる．4.2 節のアルゴリズムでは，適当な閾値を設定し輝度の高い部分のみを取り出しているが，この際，900 nm の映像の閾値は低めに設定する必要がある．本実験では，uyvy 形式のデータで y ($0 \leq y \leq 235$) の値のみを利用し，閾値と比較している．具体的な閾値は CCD の能力に依存するので，実験を通じて決定する．

第 2 に，複数の赤外画像に写っている場合には，経験的に波長の長いほうの画像をベースに点の位置の照合を行う方がよい．上記で述べたように閾値を低く設定しても，全体的に映像の雑音が少なくなるようである．これは上記と同じ理由で，CCD の感度がやや落ちるので，LED など輝度の高いものが浮き出てくるからと考えられる．

第 3 に，赤外パスフィルタが通す波長は，入射角にも影響する．カメラに取り付ける場合はあまり問題はないが，LED などに取り付ける場合には注意が必要である．なるべく照射角が狭い LED を使い，また 図 7 に示すように，反射光を防ぐために LED を黒い紙などで封印することが望ましい．

第 4 に，赤外光源の外側の点を結んでいるので，いろいろな方向から撮影しても対象物の輪郭が浮き出るように，十から数十個以上の光源を対象物に配置する必要がある．しかし，クリックするという使用目的からは厳密な領域を同定する必要はないと考える．また，アルゴリズムの関係から領域は凸となるが，対象物が必ずしも凸型であるとは限らない．しかし，これらの領域の正確な判断には形状に関する知識を与えるか，たとえば色情報などを利用して境界を同定するアルゴ

リズムを導入し，厳密さを増すことはできると考える．

これまで，赤外光を利用したイメージマップ撮影について述べてきたが，検討すべき項目もいくつかある．本方法では赤外光を使っているので，太陽光が入ると，仮に画面の一部であっても，カメラのホワイトバランスが効いて相対的に赤外光源を検出しにくくなる．また，電球やその反射光が赤外光源の背後にある場合にも光源の検出はできない．また環境の照明は，蛍光灯のような放電タイプのランプやLED照明が望ましい．このような制約を考慮しての設置が必要となる．

6. 終わりに

本論文では，実体を特別なデジタルカメラで撮影し，クリック可能としたい対象が写っている領域と，そのリンク先URLや撮影時刻など関連する情報も同時に取得/記録/再生するモデルを提案した．また，その実現方式として，赤外光の波長差分を利用して対象物の領域を同定，追跡する手法について述べた．本方式で撮影した映像をブラウザなどで提示する際に，クリック可能な対象の範囲やリンク先URLなども埋め込まれた関連情報から得ることが可能であることを実験を通して示した．また実時間処理も可能であることが分かった．

本方式は，波長差分を利用しているので，点滅させる方式（たとえば文献9)，14)）と異なり，一時的に障害物により対象物が見えなくなっても問題なく対処できる．また，光源の認識も1フレームの解析のみでできるので応答も格段に速い．映像内の点の個数に関しては実用範囲で十分にスケールするが，波長の種類が増えるとデコードすべき画像の枚数が増える．そのため，試作では光源の種類を5種としたが，これを問題と考える場合もあるかもしれない．現状では，関連情報の取得に位置依存（location dependent）な方法を用いて対処している．これでよいという意見もあるが，はじめの要求条件を少し緩くすれば，他の方法とあわせることで，格段に種類を増やせる可能性もある．たとえば，赤外線に情報を載せた変調光の利用[9),14)]，場合によっては，これらの技術にフレームレートの高いカメラをあわせることも考えられる[10),17)]．ハードウェアDVデコーダを利用することも考えられる．

前章で述べた以外の応用として，博物館などで撮影し，後に再生しクリックするとその解説がWeb経由で取得できるようになるだろう．この場合にはレーザ光の反射を使うのがよいと思われる．また，これを利用したエンタテインメントも可能かもしれない．クリックすることとは直接関係ないが，点の個数にはある程度スケールすること，および実験システムのように波長ごとの映像が独立に取り出せることから，モーションキャプチャにも利用できる．たとえば，数人に特定波長の光源を取り付け，それらをトラックすれば，波長ごとに映像が分離されるので，これまでより簡易なアルゴリズムで，複数対象物のモーションキャプチャが可能となる．実際に我々の実験システムであっても，5人までなら絡み合った動作も実時間でキャプチャが可能である．この場合にもハードウェアDVデコーダを利用すれば，全体としての処理負荷をかなり抑えられるので，光源の種類を増やすこともできる．

参考文献

1) ATSC (2004). http://www.atsc.org/
2) Azuma, R.T.: A Survey of Augmented Reality, *Presence: Teleoperators and Virtual Environments* Vol.6, No.4, pp.355–385 (1997).
3) MPEG-7 (2002). http://www.mpeg-industry.com/
4) SMIL (2001). http://www.w3.org/AudioVideo/
5) libdv. http://libdv.sourceforge.net/
6) Rekimoto, J. and Ayatsuka, Y.: CyberCode: Designing Augumented Reality Environments with Visual Tags, *Proc. DARE2000*, pp.1–10, ACM (2000).
7) Rockefeller, T.: *Convex Analysis*, Princeton Univ. Press (1970).
8) Helical-scan digital video cassette recording system using 6.35 mm magnetic tape for consumer use, IEC61834 (1998).
9) 青木 恒：カメラで読み取る赤外線タグとその応用，インタラクティブシステムとソフトウエア研究会（WISS2000），pp.131–136, 日本ソフトウエア科学会 (2000).
10) 伊藤禎宣，角 康之，間瀬健二：赤外線IDセンサを用いた設置・着用型インタラクション記録装置，インタラクション2003論文集，pp.237–238 (2003).
11) 坂田哲夫，柴垣 斉，佐藤哲司：映像散策のためのビデオハイパーメディア，マルチメディアと分散処理，Vol.97, No.13, pp.19–24, 情報処理学会 (1997).
12) 菅原俊治，青柳滋己，佐藤孝治，高田敏弘：クリッカブルオブジェクトの撮影，インタラクティブシステムとソフトウエア研究会（WISS2001），pp.167–172, 日本ソフトウエア科学会 (2001).
13) 椎尾一郎，早坂 達：モノに情報をはりつける－RFIDタグとその応用，情報処理，Vol.40, No.8, pp.846–850 (1999).
14) 田浦善弘ほか：テンプレート追跡による光学タグ認識，インタラクション2004論文集，pp.85–86

(2004).
15) 高田敏弘ほか：Cmew：連続メディアと WWW の統合，*Proc. Japan WWW Confence 97* (1997).
16) 高田敏弘ほか：応答性/対話性を重視したビデオマークアップエディタ，インタラクティブシステムとソフトウエア研究会（WISS97），pp.15–22，日本ソフトウエア科学会 (1997).
17) 松下禎宣，日原大輔，後　輝行，吉村真一，暦本純一：ID Cam：シーンと ID を同時に取得可能なイメージセンサ，インタラクション 2002 論文集，pp.9–16 (2002).

(平成 16 年 3 月 25 日受付)
(平成 17 年 3 月 1 日採録)

菅原　俊治（正会員）

1982 年早稲田大学大学院理工学研究科（数学専攻）修士課程修了．同年日本電信電話公社入社（武蔵野電気通信研究所基礎研究部）．以来，知識表現，学習，分散人工知能，マルチエージェントシステム，インターネット等の研究に従事．1992〜1993 年マサチューセッツ大学アムハースト校客員研究員．現在，NTT コミュニケーション科学基礎研究所主幹研究員．博士（工学）．日本ソフトウエア科学会，電子情報通信学会，ISOC，IEEE，ACM 各会員．

栗原　　聡（正会員）

1992 年慶應義塾大学大学院理工学研究科計算機科学専攻修士課程修了．同年日本電信電話株式会社入社．基礎研究所を経て現在，未来ねっと研究所に所属．1998 年から慶應義塾大学大学院政策・メディア研究科専任講師兼務（有期）．2004 年から大阪大学産業科学研究所知能システム科学研究部門助教授（大学院情報科学研究科情報数理学専攻兼務）．分散協調人工知能，インターネット，ネットワーク科学等の研究に従事．著書『社会基盤としての情報通信』（共立出版，共著）．博士（工学）．日本ソフトウェア科学会，人工知能学会各会員．

青柳　滋己（正会員）

1965 年生．1988 年東京工業大学理学部情報科学科卒業．1990 年 3 月同大学大学院理工学研究科情報科学専攻修士課程修了．同年日本電信電話株式会社入社．現在，同社コミュニケーション科学基礎研究所に所属．複合メディア情報処理の研究に従事．電子情報通信学会，日本ソフトウェア科学会各会員．

佐藤　孝治（正会員）

1967 年生．1989 年慶應義塾大学理工学部数理科学科卒業．1991 年同大学大学院理工学研究科計算機科学専攻修士課程修了．同年日本電信電話株式会社入社．現在，同社サイバースペース研究所に所属．分散システム，マルチメディアシステム，オペレーティングシステム等の研究に従事．日本ソフトウェア科学会会員．

高田　敏弘（正会員）

1962 年生．1986 年東京工業大学理学部情報科学科卒業．1988 年同大学大学院理工学研究科情報科学専攻修士課程修了．同年日本電信電話株式会社入社．基礎研究所，未来ねっと研究所等を経て，現在，コミュニケーション科学基礎研究所主任研究員．1994 年スタンフォード大学客員研究員．並列オブジェクト指向計算，分散システム，Web における情報発信/多言語化/マルチメディア処理，実空間コンピューティング等の研究に従事．ACM，日本ソフトウェア科学会各会員．