

## 撮影時におけるマーカを利用したタグ付けによる 映像制作支援システム

中井川 峻<sup>†</sup> 高橋 伸<sup>††</sup> 田中 二郎<sup>††</sup>

本稿では、映像制作における撮影時に2次元マーカによるタグ付けを行うことにより、タグ情報を基にした半自動的な編集を可能とするエンドユーザ向け実写映像作品制作支援システムについて述べる。映像制作の企画段階で作品に必要なシーン等の情報を入力することによりタグ情報とマーカのセットをシーンの順序に沿って生成する。撮影時に生成された順番に従ってマーカをカメラに写すことにより、シーンの順序に沿ってタグが付加された映像を生成することが出来る。そして付加されたタグ情報を利用することで企画内容に沿った自動的な編集を行うと共に、その後の手動による編集も可能とする。このように、タグとマーカを組み合わせることにより、企画、撮影、編集という映像編集の過程の結びつきを強め、エンドユーザにも容易に映像を制作することができる環境を実現する。

### Video editing support by tagging on the fly while shooting video

SHUN NAKAIGAWA,<sup>†</sup> SHIN TAKAHASHI<sup>††</sup> and JIRO TANAKA<sup>††</sup>

In this paper, we describe a live-action video production support system for end users enabling semiautomatic editing based on tag information by performing the tagging with the two-dimensional marker at the time of shooting a video in the video production. The system generates the set of the tag information and the two-dimensional marker along the order of the scene by a user inputting the information of the scenes necessary for a video at a plan stage of the video production. The system can produce the video which is tagged to along the order of the scene by shooting a marker according to a generated order at the time of shooting a video. And the system performs the automatic editing along plan by using added tag information and enables the editing by the manual operation later. In this way, we realize the environment where an end user can produce a video easily by developing ties for a plan, shooting a video and editing by combining a marker with a tag.

#### 1. はじめに

YouTube やニコニコ動画 など、動画共有サービスの流行と映像制作環境の普及により、映像作品を見るだけでなく制作して世界中の人と共有することが一般的になりつつある。動画共有サービスに投稿される作品には、楽器演奏、工作などの成果発表や技術講座をはじめとした実写映像を用いたものがあり、人気カテゴリとして挙げられる。実写映像作品はパフォーマンスの質が同程度であった場合、映像編集を効果的に利用した、映像としての質が高い作品が高く評価される傾向にある。しかし、映像編集ソフトウェアが普及

しつつある現在でも、無編集の作品も数多く見受けられる。

その原因は、編集が困難な作業であり、時間と労力がかかるためであると考えられる。より良い作品を制作するために映像の撮り直しを行うに従って映像素材ファイルが増加するが、ただでさえ検索性の低い映像というメディアから必要なファイルを探し出し、トリミングの始点や終点などの編集タイミングを見つけ出すことは非常に困難である。また、映像上で顔を隠してプライバシーを守るための合成編集や、CGの小道具を合成し、あたかもそこにあるように見せるなど、エンターテインメント性を向上させる手法としても動画内の座標情報を利用した編集へのニーズが高まっている。しかし、映像中で動く対象の座標を手動で追跡しながら編集を行うのは大変手間のかかる作業であり、映像制作への敷居を高くする要因となっていると考えられる。

<sup>†</sup> 筑波大学 大学院システム情報工学研究科  
University of Tsukuba, Graduate School of Systems  
and Information Engineering

<sup>††</sup> 筑波大学 システム情報系  
Faculty of Engineering, Information and Systems

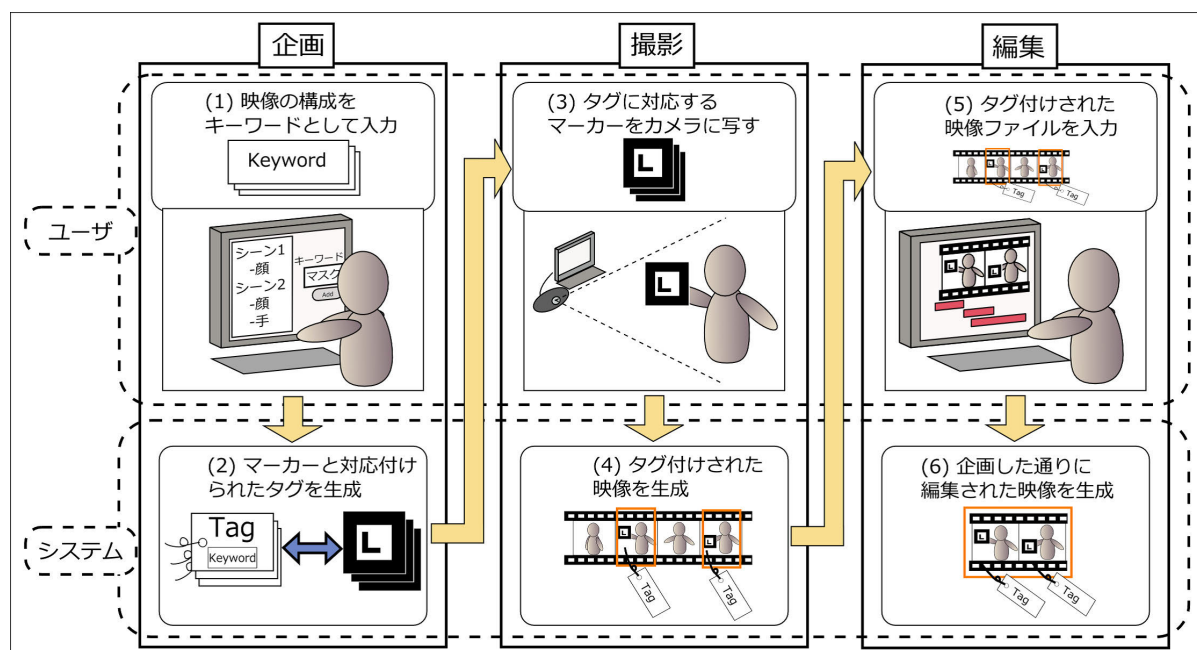


図 1 システムの全体像  
Fig.1 The system overview

上記の背景から、我々はエンドユーザ向けの映像制作支援が必要であると考えた。そこで、本研究では、(1) 企画内容のタグ化と、(2) 撮影中における 2 次元マーカーを用いた実世界でのタグ付け、(3) 付加されたタグを用いた自動編集、という 3 段階で映像制作支援を行うシステムを提案する。これにより、マーカーを順番に写しながら撮影を行うという単純な操作による、企画に沿った簡易的な映像編集が可能となる。また、マーカーの座標・向きパラメータや、タグ付けされた時間区間、タグに付加された文字列などの情報を自動編集後に利用することで、特定物体へのモザイクや画像の重畳表示といった座標・向きを用いた編集や、文字列クエリによるタイミング検索など、手動編集の支援も行う。

## 2. マーカーを用いたタグ付けによる自動編集手法

### 2.1 実写映像作品の制作プロセス

現在の実写映像制作における問題を解決するために、本研究では映像制作の過程に注目した。映像制作の過程は一般的に企画、撮影、編集という 3 つの大きなフェーズに分けられる<sup>1)</sup>。各フェーズの大まかな内容は次の通りである。

#### (1) 企画：構成を決定する

- シナリオの決定
- 必要なシーンの割り出し
- 必要な小道具の割り出し

#### (2) 撮影：カメラを用いて映像素材を撮影する

- 必要な映像をシーン単位で撮影

#### (3) 編集：映像作品として完成させる

- シーンの時間・位置的な配置
- 字幕や図、音声の合成などの高度な編集

企画フェーズではシーン等の映像構成を決定するが、映像編集ソフトはその情報を直接利用することが出来ない。そのため、編集フェーズではユーザ自身が撮影した映像素材の内容を確認しながら、企画内容と照らし合わせて編集を行う必要がある。編集フェーズにおいて自動的にメタデータを付加することで映像素材の検索性を向上させる手法<sup>2)</sup>も提案されているが、企画内容との対応付けはユーザが行う必要がある、または事前準備に専門知識や技術が必要であるため、エンドユーザには適さないと考える。

### 2.2 マーカーによるタグ付け

上で述べた実写映像作品の制作環境における問題点を考慮し、撮影した映像素材の内容確認と、映像素材と企画内容と対応付けの手間を軽減することで、より容易な動画制作環境が実現できると考えた。そこで本研究では(1) 企画フェーズにおいて、ユーザが考えた企画内容をタグと、それを操作する 2 次元マーカーの組という形で出力し、(2) 撮影フェーズにおいて、カメラにマーカーを写すだけでタグ付けを行い、(3) 編集フェーズにおいて、映像に付加されたタグに含まれる企画内容に沿った自動編集を行う、という 3 段階で映像制作を支援するシステムを提案する。提案システ

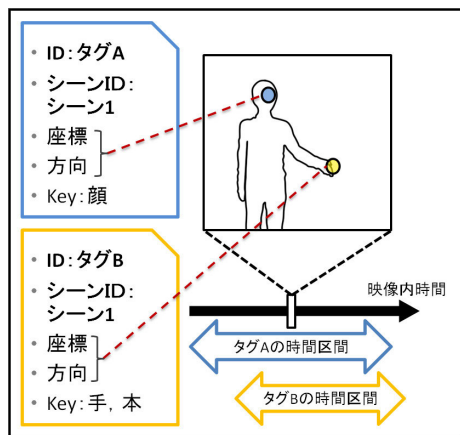


図 2 タグに含まれる情報

Fig. 2 The information in a tag

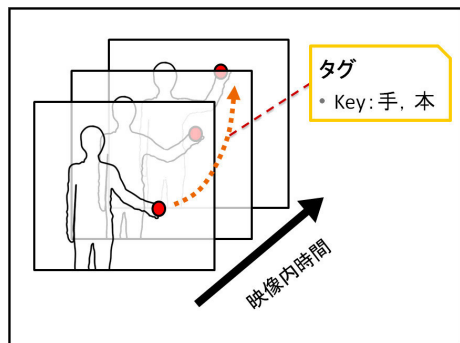


図 3 時間遷移するタグの座標

Fig. 3 The coordinate of a tag changing at time

ムの全体像を 図 1 に示す。

ここで、本研究における「タグ」とは、図 2 のように撮影した映像素材ファイルの特定の時間区間に対して紐付けることができるメタデータである。タグの内容は、(a) タグ付けされた時間区間がどのシーンであるかを表すシーン情報、(b) ユーザが入力したキーワード文字列、(c) タグの 3 次元座標・向き情報とした。システムは (a) を利用することで、ユーザが定義したシーンとタグ付けされた時間区間の対応付けを行うことが出来る。また、(b) により、キーワードと映像内の時刻を対応づけることが出来る。さらに、(c) により、図 3 のような、時刻に対応したオブジェクトの動きを扱うことが出来る。

## 2.3 ビデオ制作の流れ

### 2.3.1 企画フェーズ

企画フェーズでは、ユーザはあらかじめ動画のシナリオや構成を考えて、必要なシーンと、シーン毎に必要な小道具などの「オブジェクト」を割り出す。そして、そのシーンやオブジェクト各々のキーワードを決めて企画モジュールに入力する(図 1(1))。すると、企

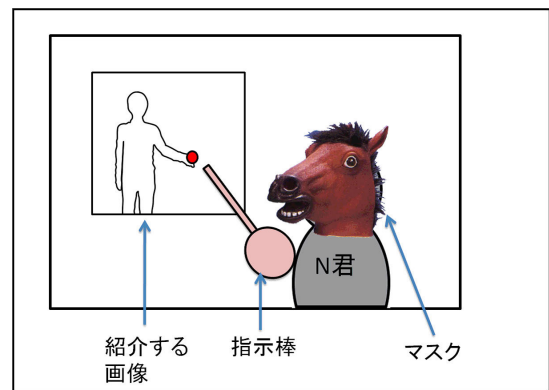


図 4 研究紹介映像のイメージ図

Fig. 4 Image of introductory video of this research.

画モジュールは各キーワードに対応するタグを内部で生成する。さらに、ユーザが撮影フェーズで用いるマーカーを各タグに対して出力する。なお、シーンのタグに対しては「開始」「終了」という 2 つのマーカーを、またオブジェクトのタグに対しては 1 つのマーカーを出力する。

ここでは例として、本研究を紹介するデモビデオを作成することを想定して、具体的に説明する。ビデオの登場人物は N 君であり、図 4 に示すイメージ図の様に、N 君が本システムにおける企画、撮影、編集について、関係する画像を指示棒で指し示しながら説明を行う、とする。シーン構成は以下の 3 シーンとする。

- (1) 「オープニングシーン」最初に始めの挨拶をする
- (2) 「解説シーン」本システムについて解説を行う
- (3) 「エンディングシーン」最後に終わりの挨拶をする

また、撮影後に以下のような合成処理を行いたいとする。

- N 君は顔を出すのが恥ずかしいので、全てのシーンで顔を「マスク」のオブジェクトで隠すようにする。
- 解説に用いる画像は、撮影後に合成してビデオに埋め込む。また、埋め込み画像をうまく指示する演技は難しそうなので、解説シーンでは「指示棒」オブジェクトを用いて画像中の解説したい位置を指示するように編集を行うことにする。

これでビデオの企画が固まったので、キーワードを入力する。具体的には、3 つのシーンのキーワード「オープニング」「解説」「エンディング」と、解説シーンで用いる指示棒のオブジェクト用の「指示棒」、全シーンで顔を隠すマスク用の「マスク」という 2 つのオブジェクトのキーワードをそれぞれ入力する。すると、シーン/オブジェクトのタグがそれぞれ生成され、

またそれぞれに対しマーカーが出力される。例えば、オープニングシーンについては「オープニングシーン開始」マーカー、「オープニングシーン終了」マーカーという2つのマーカーが出力され、指示棒オブジェクトには「指示棒オブジェクト」のマーカーが出力される。

### 2.3.2 撮影フェーズ

撮影フェーズでは、ユーザは企画内容に沿ったシーンの撮影を行いながら、撮影モジュールのカメラにマーカーを写すことでタグ付けを行う。ユーザはマーカー画像を印刷して手に持ったり頭部付近に貼りつけたり、マーカー画像の表示とその切り替えを行うための携帯端末向け撮影操作アプリケーションを利用して任意のタイミングで任意のマーカーをカメラに写すことが出来る(図1(3))。シーントグを付加する場合は、シーンの開始としたいタイミングでシーン開始マーカーを写し、シーンを終了させたいタイミングでシーン終了マーカーを写す。これにより、シーン開始マーカーが写された時間区間の終点と、シーン終了マーカーが写された時間区間の始点の間の時間区間にシーントグを付加することが出来る。また、オブジェクトタグのマーカーを写し続けることで、写していた時間区間にオブジェクトタグを付加することが出来る。撮影を終えると、タグ付けされた映像をプロジェクトファイルとして出力する(図1(4))。

具体的には、マスクオブジェクトタグを頭部付近に付加したいため、その付近にマーカーを付けることにする。ユーザは「マスク」オブジェクトのマーカーを紙に印刷して首の部分に装着する。指示棒オブジェクトタグやシーントグの操作には撮影操作アプリケーションを用いることにする。

次に、撮影モジュールを起動して、ディスプレイに表示されるプレビュー画面を見ながら撮影範囲を調整する。して撮影を開始する。撮影操作アプリではシーンリストから撮影するシーンをタッチしてマーカーセットを表示させ、最初のマーカーである「オープニングシーン開始」を選択することでマーカー画像を拡大表示する。これをカメラに向けて写した後、それを隠した瞬間からオープニングシーントグが映像に付加され始める。同時に首に付けたマスクのマーカーも認識されており、マーカー位置にマスクのタグが付加される。オープニングシーンの撮影を行った後、シーンを終了する際には撮影操作アプリケーション上に表示されたマーカー画像を左へスワイプ操作することにより、「オープニングシーン終了」マーカーを表示させ、それをカメラに写した瞬間にオープニングシーントグ

の付加が完了する。解説シーンではシーントグの付加操作に加え、「指示棒」オブジェクトのタグを表示させた端末を手に持ち、説明している様な仕草をすることにより、あたかも本当に指示棒を動かしているようにタグ付けを行う。エンディングシーンに関してはオープニングとほぼ同様の流れでタグを付加しながらシナリオに沿った撮影を行う。このように、基本的には企画モジュールで作成した順番にマーカーを写すだけで必要なタグ情報を付加することが出来る。

### 2.3.3 編集フェーズ

編集フェーズでは、ユーザは編集モジュール(映像編集ソフトウェア)に撮影モジュールから出力されるプロジェクトファイルを読み込ませる(図1(5))。すると、シーントグが付けられた部分の映像のみを切りだし(トリミング)、企画モジュールにキーワードとして入力された順番に並び変えた状態にプロジェクトが編集される(図1(6))。また、タグ付けした時間区間には、そのタグに付加されたキーワード情報が付加されると共に、マーカーの座標と向きを反映した平面型のオブジェクトを付加した状態に編集される。ここでの編集は編集ソフトウェアのインタフェース上で行われるものであり、映像ファイルとして出力されるわけではない。ユーザは、システムによって大まかに自動編集されたプロジェクトを、編集モジュールのインタフェースを操作して修正することが可能である。例えば、ユーザは「顔」という文字列のクエリで検索することで、タグ付けされた時間区間の始点を探してタグの座標、つまり顔の座標にCG画像を合成したり、「オープニング」という文字列のクエリで挨拶のシーンを検索し、その開始時間を前後に調整した後に映像ファイルとして出力する、といったことが可能である。

研究紹介デモビデオの制作においては、プロジェクトファイルを編集モジュールに読み込ませることで、オープニングシーンのタグが付加された時間区間、解説シーンのタグが付加された時間区間、エンディングシーンのタグが付加された時間区間、が編集モジュールのインタフェース上に順番に並び、シーントグが付加されていない部分は自動的にトリミングされる。さらに、マスクのオブジェクトタグの座標と向きに従い、各シーンにおいてキーワードと共に平面オブジェクトが追加される。自動編集を行った結果は図5のようになる。平面オブジェクトの座標と向き(回転量)は、その後にCGを重ね合わせる用途などに利用でき、イメージした様な映像を完成させることが可能である。

このように、本手法を用いることにより、(a)企画内容をリアルタイムにタグ付け可能となり、素材の内

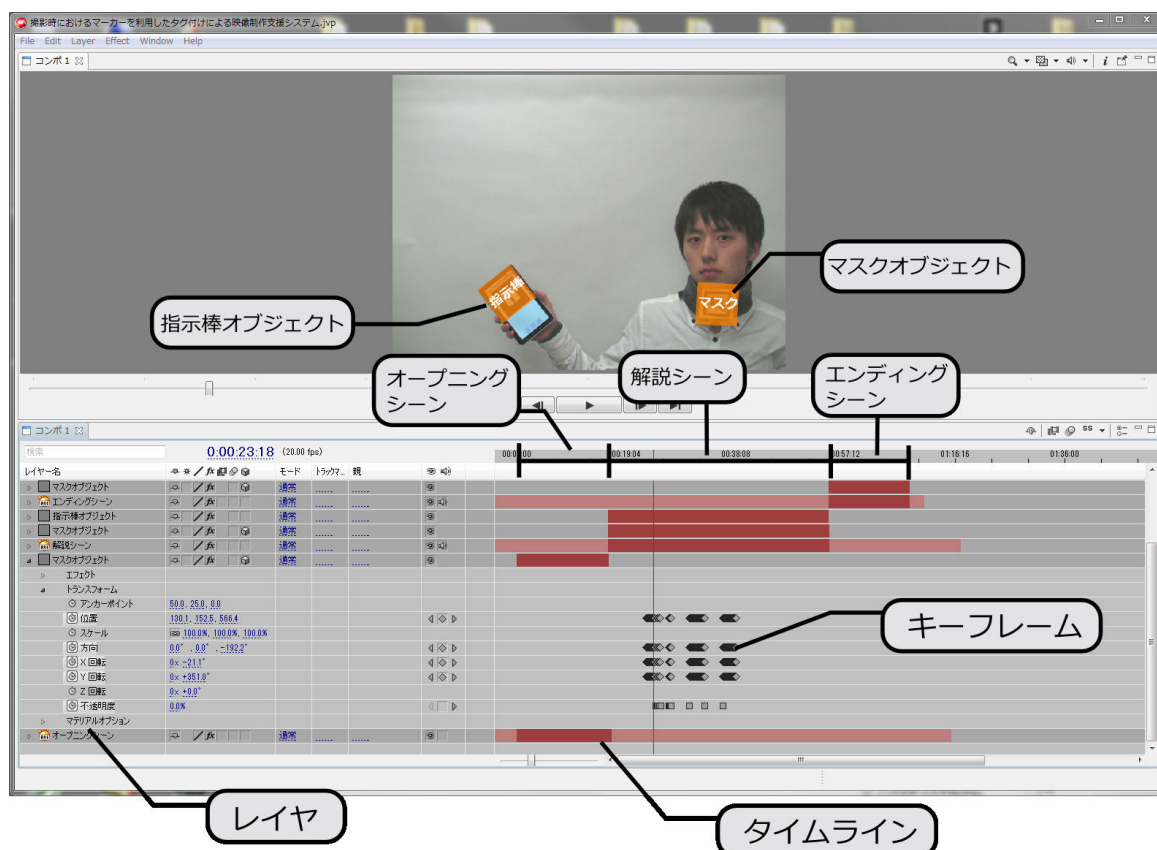


図 5 編集モジュールでの自動編集結果

Fig.5 Automatic editing result in the editing module

容確認や企画内容との対応付けの手間を削減できる，(b) 被撮影者自身が操作できるため，追加人員が必要ない，(c) 実世界での容易な操作で自動編集した映像を生成出来るため，エンドユーザに優しいだけではなく，作業時間の削減につながる，(d) 撮影されたマーカーの様子から 3 次元座標・向きを取得することが出来るため，座標・向き情報を利用した CG 合成などの高度な編集を支援することができる，といったメリットが得られると考えられる．

### 3. システムの実装

提案システムは企画モジュール，撮影モジュール，編集モジュールという 3 つのモジュールから構成され，それぞれのモジュールが連携することでタグ情報を介した映像制作を実現する．それぞれの実装の詳細について以下に述べる．

#### 3.1 企画モジュール

企画モジュールは入力されたシーンやオブジェクトに関するキーワードを基に順序立てたタグ群を生成する．タグにはどのシーンに所属するかを表すシーン情報，複数のキーワードと，タグに対応する 2 次元

マーカーの ID 情報が含まれる．2 次元マーカーは事前に用意した ARToolKit<sup>3)</sup> のマーカーファイルからランダムで選択し，タグに対応付ける．最後に，これらのタグの順序や含まれるキーワード情報，シーン情報や，タグとマーカーの対応関係情報を含むプロジェクトファイルを作製し，ネットワークを介して他のモジュールと共有する．

#### 3.2 撮影モジュール

撮影モジュールは Web カメラから取得した画像と音声を圧縮しながら録画する．また，録画と同時にカメラ画像に対して ARToolKit を用いた 2 次元マーカー認識を行い，ARToolKit から得られるマーカーの座標変換行列を 1 フレーム毎に記録していく．このように，マーカー情報を映像と同時に記録することで，映像へのタグ付けを実現している．撮影操作アプリは Android 端末上に実装した．撮影操作アプリでは，マーカー画像とタグ情報を含むプロジェクトファイルを，DropboxAPI によりネットワークを介して企画モジュールと共有する．これを読み込むことにより，タグに関連付けられたマーカー画像から，任意のマーカー画像一枚を表示することが出来る．



### 3.3 編集モジュール

編集モジュールは企画フェーズで作成したタグの情報と、撮影フェーズで作成した動画情報を読み込ませることにより映像の自動編集を行う。実装は Javie<sup>4)</sup>というオープンソース映像編集ソフトウェアに機能を追加する形で行った。システムの概観は図5に示した通りである。JavieはAdobe AfterEffectsに似たインタフェースを持つ映像編集ソフトウェアであり、映像ファイルや静止画ファイルなどをレイヤとして扱うことができる。また、レイヤやレイヤに付加したエフェクトが持つパラメータを、キーフレーム単位で制御することが出来る。編集モジュールは、撮影時に認識されたマーカの情報から、シーンタグの付加された区間を求め、プロジェクトに追加した映像レイヤをシーンタグの付加された区間に合わせてトリミングするしつつ、タグを付加する。この時付加されるタグはレイヤに追加するエフェクトとして実装している。映像レイヤにタグエフェクトを適用することにより、キーワード文字列を利用した文字列クエリを用いた時間区間検索を行うことが出来る。文字列クエリによる検索を行う際には、このテキストパラメータが設定されている最初のタイミングを探索し、探索したタイミングまでプレビュー映像をシークさせている。オブジェクトタグのマーカが認識された区間には、単色の平面型レイヤを追加している。平面レイヤには3次元座標と3次元の向きを設定することが可能であり、オブジェクトタグのマーカの各フレームにおける座標変換行列から求めた座標、回転角をキーフレームに設定することにより、マーカの動きを再現する。これにより、手動編集時にタグ付け位置の視覚的な確認を可能とするとともに、他のレイヤの座標、回転を座標軸、回転軸として設定することで、タグの座標を利用した編集を可能とする。

### 4. 関連研究

動画編集のための自動的な索引付けを行う研究として、Kumanoらの研究が挙げられる<sup>2)</sup>。この手法では、映像文法という独自に定義した映像のルールに基づき、素材として利用できる区間などの映像の状態遷移を自動認識し、検索を行うための索引を付加する。これに対して本研究の提案システムでは、撮影前の準備がキーワード入力のみであり、また、編集モジュールに映像を読み込んだ時点で索引が付加された上で編集された状態となる。これらの点から、エンドユーザにとってより容易に映像制作を行える環境であると言える。

SmithらのVideo-Skims<sup>5)</sup>では、まず、映像内の音声から作成したトランスクリプトからTF-IDF法を用いて重要なキーワードを抽出する。そして、顔検出やカメラワークの検出などを手がかりにしてキーワードを発話した時間区間を認識し、認識した区間情報を利用して要約映像を制作することが出来る。これに対して本研究の提案システムでは、自身で定義した企画内容を基にしたタグを任意の区間に自由に付加できるため、意図した構成の映像を制作することが出来る点で異なっている。

### 5. まとめ

エンドユーザ向けの映像制作を支援する新しいシステムを提案し、プロトタイプの実装を行った。本システムは、企画時に入力した情報を基に生成したタグをマーカと対応づけることにより、マーカを使った撮影中のタグ付けによる半自動的な映像作品編集を可能とする。今後はエンドユーザを対象としてシステムを実際に利用してもらい、評価を行う予定である。

### 参考文献

- 1) 岡村征夫：映像編集入門，オーム社（2010）。
- 2) Kumano, M. and Ariki, Y. : Automatic Useful Shot Extraction for a Video Editing Support System, in Proc. MVA, 2002, pp.310-313.
- 3) Kato, H. and Billinghamurst, M. : Marker tracking and HMD calibration for a video-based augmented reality conferencing system, Proc. 2nd IEEE and ACM International Workshop on Augmented Reality, pp. 85-94, (1999).
- 4) Javie .<http://sourceforge.jp/projects/javie/> (accessed 2011-12) .
- 5) Smith, M. and Kanade, T. : Video Skimming and Characterization through the Combination of Image and Language Understanding, Proc. ICCV98, pp. 61-70, (1998).