

カメラワーク情報を主なパラメータとする 映像オーサリングシステムに関する検討

高木真一†* 池上大介‡ 浜崇之‡ 石川孝明‡ 石川裕也‡
近藤雅恵‡ 松岡正悟‡ 小舘亮之* 富永英義†‡*

†早稲田大学大学院国際情報通信研究科

* 早稲田大学国際情報通信研究センター

〒169-0051 東京都新宿区西早稲田 1-3-10 早稲田大学 29-7 号館

Tel: (03)5286-3831

E-mail: takagi@tom.comm.waseda.ac.jp

‡早稲田大学理工学部電子・情報通信学科

〒169-8555 東京都新宿区大久保 3-4-1

Tel: (03)5286-3385

あらまし 本稿では、一般個人が撮影した映像を対象とし、映像より抽出したカメラワーク情報を主なパラメータとする映像オーサリングシステムに関する検討を行ったので報告する。まず、一般個人によって撮影された映像の特徴とダイジェスト映像作成時の編集作業について考察を行い、映像編集におけるカメラワーク情報の有効性を考察し、また映像より抽出されたカメラワーク情報を利用して、ショットをさらに細かくセグメント化し、セグメントごとのカメラワーク安定性を自動評価するアルゴリズムの検討を行った。

キーワード 映像編集 カメラワーク 動きベクトル セグメント分割

A Study on Method to Analyze Camerawork Information for Video Authoring System

Shin'ichi TAKAGI†* Daisuke IkeGAMI‡ Takayuki HAMA‡
Takaaki ISHIKAWA‡ Yuya ISHIKAWA‡ Masae KONDO‡ Shogo MATSUOKA‡
Akihisa KODATE* Hideyoshi TOMINAGA†‡*

† GITS, WASEDA University

* GITI, WASEDA University

Bldg.29-7,1-3-10 Nishi Waseda, Shinjuku-ku, Tokyo, 169-0051 Japan

Tel: +81-3-5286-3831

E-mail: takagi@tom.comm.waseda.ac.jp

‡Dept. of Electronics Information and Communication Engineering, Waseda University

3-4-1 Okubo, Shinjuku-ku, Tokyo, 169-8555 Japan

Tel: +81-3-5286-3385

Abstract In this paper, we have studied a video authoring system especially for non-professional person, which mainly uses camerawork information which was extracted from video stream. After we discussed a effectiveness of camerawork information in the video authoring systems, we proposed a new algorithm to divide a video shot into segment using camerawork information. This algorithm also includes a method to evaluate a camerawork stability of input video. This means that automatic video editing system could be realised using this algorithm.

key words Video Editing Camerawork Motion Vector Segmentation

1. はじめに

コンピュータ関連技術の進展により、以前は高度規格の専用機材を用いていたような映像編集作業が、デジタルビデオカメラやPC上の安価なノンリニア映像編集ソフトの登場・普及によって、一般個人でも特別な専用機材を用いることなくできるようになりつつある。また、xDSLやFTTHの普及により、一般個人でもインターネット上の数Mbpsの高品質映像のストリーミング視聴が現実になってきており、自ら高品質映像を配信することも容易になりつつある。これらのことは、従来、専門知識・専門装置を持った限定された人にしかできなかった映像コンテンツの制作・配信作業の一部に、一般個人が関与できる方向へ向かっていると考えることができる[1]。しかし、映像編集ソフトが以前と比較して非常に安価に入手可能であるものの、一般個人がビデオカメラで撮影した映像素材から映像コンテンツを制作するには、映像編集の知識や経験が不足していたり、その作業自体が多く時間を必要とするため、簡単に短時間でできる作業とは言えず、ビデオカメラで撮影した映像素材を編集しないまま放置していることが多い。そのため、一般個人でも簡単に時間をかけずに映像コンテンツ制作ができるように支援する技術やシステムが求められている。

我々の研究グループでは、一般個人が市販のデジタルビデオカメラによって撮影した映像素材から、そのダイジェスト映像を制作する作業の手間を大幅に削減、または自動化させることを目的とした、映像編集システムを検討している[2]。本稿では、主にカメラパラメータを使った映像セグメント分割法とカメラワーク安定度の評価アルゴリズムについて検討を行ったので報告する。

本稿は、まず2章で一般個人による撮影と編集作業に関する考察を行い、3章で今回検討したシステムの概要について示し、4章ではシステムで扱うカメラワークを整理する。5章では、カメラワーク情報抽出手法とその情報を利用したセグメント分割法およびカメラワーク安定度評価法について、そのアルゴリズムを述べる。6章では、アルゴリズムの処理速度に関する簡易実験を行い、考察を述べる。7章では、まとめと今後の課題を述べる。

また、本稿では、映像シーケンスにおけるカット点(映像切り替え点)から次のカット点までの間をショット(Shot)と呼ぶことにする。

2. 一般個人による撮影とダイジェスト映像制作作業に関する考察

2.1 撮影映像に関する考察

一般個人が市販ビデオカメラを使って撮影した場合、撮影方法に対する知識・経験がないた

め、カメラ操作が安定しない場合が多く、激しい手ぶれが含まれるショットが比較的多い。また、ビデオカメラの電源の切り忘れによって画面ばかり映っているなど、ダイジェスト映像には意味のなさないショットも存在する。また、旅行時の映像や子供の映像を撮影することが多いが、そのような場合は映画やドラマなど撮影のように台本がある場合と異なり、あらかじめ撮影方法が事前に決定されているわけではなく、カメラの前での出来事を、興味の向くままに次々と撮影していくことが多いため、一つのショットの中に複数のカメラワークが混在することが多く、また同一ショット内で特に関連性のない映像が複数含まれていることがある。

2.2 ダイジェスト制作作業に関する考察

特別な知識や経験を持たない一般個人によるダイジェスト映像制作作業について、撮影した素材全てを編集対象とすることを想定した場合、編集作業の前に素材映像シーケンスを一度全て再生し、素材映像シーケンスのどの時間位置にどのような映像が記録されているかを把握する作業(内容把握作業)を行うのが一般的である。この作業には、素材映像シーケンスの時間長と同程度かそれ以上の時間を費やすことが多く、一般個人に対する作業負担の増加の一因となっていると考えられる。

また、内容把握後にダイジェスト映像にどの部分を採用するかを決定する必要があるが、一つのショットは数秒から十数秒である多いため、選択すべき部分が多数になり、趣味として作業を行うといった場合以外では、その作業自体が負担になることが多い。

3. システム概要

前章での考察を踏まえ、一般個人が撮影した映像を対象とし、そのダイジェスト映像制作を自動化あるいは支援し、その作業労力を軽減させることを目的として、カメラワーク情報を主なパラメータとする映像のセグメント化とカメラワーク安定度評価値決定法について検討を行った。本システムで対象とする主な映像は、市販のデジタルビデオカメラで無計画に撮影したような同一ショットに複数のカメラワークを含み、カメラ操作の安定性の低い部分を含む映像とする。本システムにおける処理の流れを図1に示し、その概要は次のとおりである。

まず、映像シーケンスをショット単位に分割をする。次に、同一ショット内の映像からカメラワーク情報を抽出し、そのカメラワーク情報をもとに、カメラワークごとのセグメントに分割する。その際に、セグメントごとにカメラワークの安定性評価値を算出する。その後、カメラワークの安定性の評価値(以下カメラワーク評価値)を利用して、自動的にセグメントを

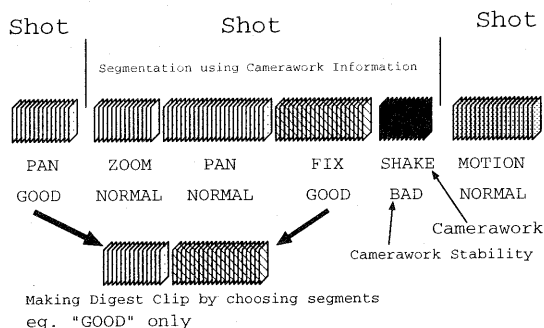


図1 システム概要

結合し、ダイジェスト映像を生成する。一連の処理は、全て自動的に行うことが可能である。

また、ユーザ操作画面には、各ショットの代表画像と、各セグメントの代表画像と時間情報、カメラワーク評価値等が表示される。ユーザが好みのダイジェスト映像を制作する時は、カメラワーク評価値を参考にしながら、必要なセグメントを選択することも可能である。また、システムに自動的に選択されたセグメントをベースをして、ユーザが好みのセグメントと自由に入れ替えることもできる。但し、本稿では、紙面の都合上、映像からのカメラワーク情報抽出方法とその情報を利用したセグメント分割方法とカメラワーク安定度評価決定法について述べる。

4. 抽出するカメラワーク

カメラ操作には主に FIX, PAN(Left, Right), TILT(Up, Down), ZOOM(In, Out), DOLLY(Forward, Backward), TRACK(Left, Right), BOOM(Up, Down), ROLL が存在する。しかし、PAN と TRACK, TILT と BOOM, ZOOM と DOLLY はそれぞれ似たような映像になることから、本システムにおいては区別しないものとした。但し、PAN と TILT に関しては、用語としては PAN に統一した。また、曖昧なカメラワークとして MOTION を定義した。上記以外に一般個人の映像において多発する手ぶれ (SHAKE) を抽出する。従って、本システムで抽出するカメラワークは、次のとおりとした。

- FIX
- PAN(Left, Right, Up, Down)
- ZOOM(In, Out)
- MOTION

● SHAKE

これらのうち、MOTION と SHAKE については、ショット内のセグメントを決定する際に抽出されるものである (後述)。

5. カメラワーク抽出とセグメント分割およびカメラワーク安定度評価値決定法

本システムにおいてカメラワーク抽出は、対象フレームを小領域に分割し、テンプレートマッチングにより小領域ごとの動きベクトルを算出することをベースとして実現する。しかし、動きベクトルの算出は、そのフレームあたりの算出ベクトル数を多くし、またその算出精度を高く設定すると、処理に多くの時間を必要とする [3]。

本システムでは、映像ストリームの内容把握を短時間で行うことが目的の一つであり、カメラパラメータ FIX, PAN(Left, Right, Up, Down), ZOOM(In, Out), SHAKE のみを区別できれば良いことを考慮し、動きベクトル算出アルゴリズムを簡易化し、また算出された動きベクトル情報を利用して、算出対象フレーム数を決定することで、カメラワーク抽出を高速化するアルゴリズムを検討した。

本アルゴリズムは、主にフレーム間処理と対象フレームにおけるカメラワーク抽出処理、セグメントおよびカメラワーク安定度評価値決定処理の3つの処理から構成される。

5.1 フレーム間処理

ここでは、テンプレートマッチングによる動きベクトル処理を全ての隣接フレーム間で行うことによる処理コスト増大を回避するために、動きベクトル算出処理対象を絞り込む処理を行う。処理の流れは次のようになる。

手順1 映像シーケンスの n 番目のフレームに対して、 $n + INT_1$ 番目のフレームを参照し、動きベクトル発生率を算出する。発生率が閾値 Th_{fix} 以下であれば、その間のフレームのカメラワークは **FIX** と判定し処理を終了し、 $n + INT_1 + 1$ 番目のフレームに対して手順1から行う。閾値以上であれば、手順2へ進む。

手順2 n 番目のフレームに対して、 $n + INT_2$ 番目のフレームを参照し、手順1と同様に動きベクトル発生率を算出する。発生率が閾値 Th_{fix} 以下であればカメラワークを **FIX** と判定し処理を終了し、 $n + INT_2 + 1$ 番目のフレームに対して手順1から行う。閾値以上であれば、 $ZOOM_1$ であるかを

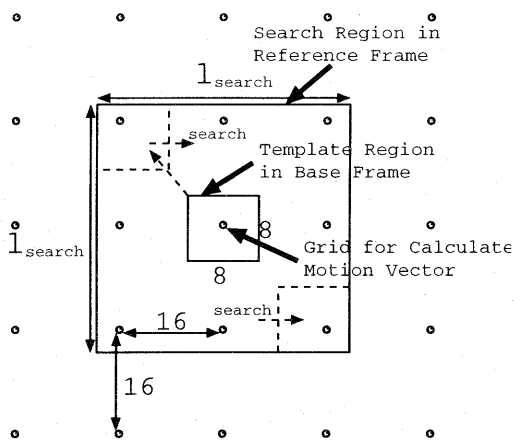


図2 動きベクトル算出領域

判定する。ZOOM₁であれば、処理を終了し、FIXの場合と同様に $n + INT_2 + 1$ 番目のフレームに対して手順1から行う。FIX, ZOOM₁のいずれにも判定されなかった場合は、手順3へ進む。

手順3 n 番目のフレームに対して、隣接するフレームごとに動きベクトルを算出し、PAN₁, PAN₂, ZOOM₁, ZOOM₂, otherの各カメラパラメータを判定する。

5.2 対象フレームにおける動きベクトル算出および動きベクトル発生率

ここでは、対象フレーム間の動きベクトルの算出方法とフレーム間処理におけるカメラワークFIXの判定に利用する動きベクトル発生率の算出方法について述べる。

動きベクトルの算出

x, y 方向とも16画素おきの画素の格子点を、ベクトル算出格子点と定義する。そして、ベクトル算出格子点を中心とした大きさ 8×8 画素のテンプレート領域を、参照フレームの探索領域において、左上から x, y 方向に m 画素ずつ平行移動させながら各位置においてその画素値の差分の絶対値を計算し、その領域内で最も差分値の小くなる時のテンプレートの中心座標位置に対するベクトル算出格子点の座標位置からのベクトルをその格子点の動きベクトルとする。探索領域は、ベクトル算出格子点を中心とした $l_{search} \times l_{search}$ の領域とする。その様子を図2に示す。本システムでは、処理速度を向上させるため、 $m = 3$ とし計算コストを削減している。

動きベクトル発生率の算出

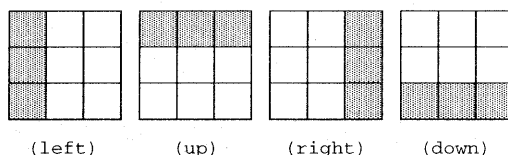


図3 方向領域

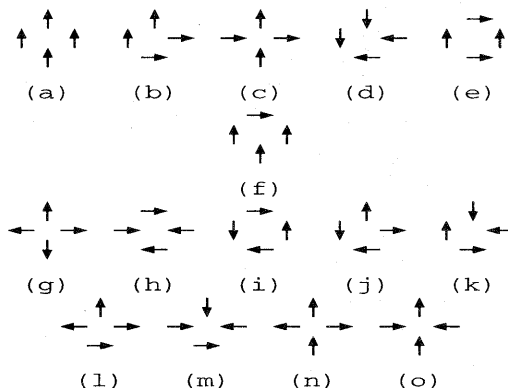


図4 方向ベクトルによるカメラワーク判定

上記で求めた各ベクトル算出格子点ごとの動きベクトルの大きさにより、発生率の計算を行う。まず、各ベクトル算出格子点を中心として、 $l \times l$ 画素の領域をベクトル発生判定領域と定義する。そのフレームにおける動きベクトル発生率 P_{vector} は、ベクトル発生判定領域の外側を指しているベクトル算出格子点の総数を $Grid_{out}$ とし、全ベクトル算出格子点数を $Grid_{total}$ とすると次の式で表すことができる。

$$P_{vector} = \frac{Grid_{out}}{Grid_{total}} \times 100$$

5.3 動きベクトル情報からのカメラパラメータ判定

ここでは、算出された動きベクトル情報を用いてカメラワークを判定するアルゴリズムを述べる。大まかな手順は次のとおりである。

手順1 フレームを9分割し、図3における各灰色領域のように左側領域(left)、上側領域(up)、右側領域(right)、下側領域(down)の各方向領域に含まれるベクトル算出格子点の動きベクトルを解析し、各領域ごとを代表する方向ベクトルを決定する。

手順2 手順1で決定された4領域の方向ベクトルの組み合わせでPAN₁, PAN₂, ZOOM₁, ZOOM₂, otherを決定する。

方向ベクトルの決定

まず、各動きベクトルを上下左右の4方向の最も近い方向に変換する。そして、図3におけるleft, up, right, downの各方向領域に含まれる4方向に変換された各々の動きベクトルで、最も多く含まれる方向の動きベクトルをその領域の方向ベクトルとする。最も多く含まれる方向ベクトルが複数方向存在する時は、変換前の動きベクトル大きさの総和を計算し、総和の最も大きいベクトルを方向ベクトルとする。

方向ベクトルからのカメラワーク判定

方向ベクトルの組み合わせは、各方向領域で上下左右の4方向存在し、上下左右の方向領域があるので、 $4^4 = 256$ 通りの組み合わせが存在する。ここでは、4つの方向ベクトルの組み合わせからカメラワークを判定する。

基本的に、方向ベクトルから推測して比較的是っきりしたカメラワークを種類ごとに PAN_1 , $ZOOM_1$ とする。 PAN_1 , $ZOOM_1$ に類似しているが、はっきりしていないカメラワークをそれぞれ PAN_2 , $ZOOM_2$ とする。 PAN と $ZOOM$ が混在しているようなものを、 ZP とし、それ以外のはっきりしないものを $other$ とする。詳細は次の通りである。

PAN_1 図4(a)のように、4つの方向ベクトルの方向が揃っているもの(4方向で4通り)。(b),(d)のようにある方向領域の方向ベクトルに対して両隣のベクトルが異なり、そのうち一方のベクトルが共通で、もう一方のベクトルが 90° 回転しているもの(8通り)。(c),(e)のように、ある方向領域の方向ベクトルに対して向かい側の方向ベクトルが同一で、両隣の方向ベクトルが 90° 回転しており、同一であるもの(8通り)

PAN_2 (f)のように(a)と比較し、4つの方向ベクトルのうち1つが 90° 回転しているもの(32通り)。

$ZOOM_1$ (g)のようなはっきりとした $ZOOM$ (2通り)。(h)のように、ある方向領域の方向ベクトルが片隣と同一で、残りの2つが 180° 回転したもの(8通り)。(i)と全体を 90° 回転させたもの(2通り)。(j)と(k)をそれぞれ全体を 90° ずつ回転させたもの(8通り)。

$ZOOM_2$ (l),(m)のように、(g)と比較して一つの方向ベクトルが 90° 回転しているもの(16通り)。

ZP (n),(o)のように、(g)と比較して一つの方向ベクトルが 180° 回転しているもの(8通り)。

$other$ 上記以外のもの全て。

上記の分類によって PAN_1 , PAN_2 , $ZOOM_1$, $ZOOM_2$ として判定されたものは、方向ベクトルの組み合わせから判断して PAN に関しては上下左右の方向が、 $ZOOM$ に関してはInかOutの方向が決定される。

5.4 セグメント決定およびカメラワーク安定性評価値の決定

ここでは、上記によって決定されたカメラワーク情報を利用して、各ショットをさらに細かいセグメントに分割する際のセグメントの決定方法およびそのセグメントにおけるカメラワーク安定性評価値の決定アルゴリズムについて述べる。

本アルゴリズムの概要は、各ショットにおいて先にはっきりしているカメラワーク部分を利用して、仮セグメントを決定し、仮セグメントに当てはまらない曖昧なカメラワーク部分の処理を行い、最終的なセグメントの構成を決定する。カメラワーク安定性評価値は、各カメラワークに対して、GOOD, NORMAL, BADの3種類のいずれかに決定される。

手順1 ショット内において、5.1の手順1,2で決定された INT_1 または INT_2 フレーム分の FIX と $ZOOM_1$ に関して、同じカメラワークが隣接していれば、結合していき、 $FIX(GOOD)$ または $ZOOM(GOOD)$ の仮セグメントを構成し、仮評価値を決定する。表記の括弧内は評価値を示す。手順1が終了した段階では、仮セグメントを構成していないのは隣接するフレームごとカメラパラメータを判定したフレームであり、非セグメントフレームは仮セグメントに挟まれた状態になっている。

手順2 非セグメントフレームに注目し、5フレーム単位でユニットを構成し、ユニットごとのカメラワークと仮評価値を決定する。具体的には、5フレームのうち PAN_1 , $ZOOM_1$ のいずれかが3つ以上占めれば、そのユニットは PAN または $ZOOM$ と判定される。また、 PAN_1 , PAN_2 , ZP のみから構成されていれば、 PAN と判定し、 $ZOOM_1$, $ZOOM_2$, ZP のみから構成されていれば、 $ZOOM$ と判定する。それ以外は、 $other$ と判定する。また、評価値に関しては、5フレームのカメラワークの方向が全て同一であればGOODと仮評価し、それ以外は、NORMALと仮評価する。

手順3 手順2において、GOODと仮評価されたユニットが同一カメラワークで Th_{good} 個以上連続する場合は、仮セグメントを構成し、 $PAN(GOOD)$, $ZOOM(GOOD)$ と仮評価値を決定する。

手順4 各ユニットを手順1で仮決定されているセグメントに統合するか、独立したセグメントとするかを決定する。手順3が完了した段階で、仮セグメント間のセグメント化されていないユニットが Th_{merge} 個以上連続する場合には、統合処理は行わずに独立した仮セグメントを構成する。独立した仮セグメント内において、 Th_{bad} 以上連続して other が存在する場合は、その仮セグメントは、SHAKE(BAD)と決定し、それ以外を MOTION(NORMAL)と決定する。仮セグメント間に Th_{merge} 以下しかユニットが存在しない場合は、手順5へ進み前後のいずれかのセグメントへの統合処理が行われる。

手順5 対象ユニットの前後の仮セグメントを比較し、前後とも同一カメラワークであれば、前後仮セグメントと対象ユニットを統合して、新しい仮セグメントを構成する。前後が異なる場合は、ZOOM,PAN,FIXの順の優先順位でユニットを統合する。統合対象となった仮セグメントの仮評価値がGOODの場合は、新しい評価値をNORMALとする。仮評価値がGOOD以外の場合には、統合処理をしても評価値は変化しない。また、統合処理をしてもセグメントのカメラワークは変化しない。

手順6 セグメント化されていないユニットがなくなるまで、統合処理をおこなう。全てセグメント化されたら、仮カメラワークと仮評価値を最終的なカメラワークとカメラワーク安定度評価値とする。

6. 処理時間に関する簡易実験

本稿で検討したアルゴリズムをコンピュータに実装し、処理速度等の簡易実験を行った。実装したコンピュータのスペックは、CPUがIntel Xeon 2.2GHz、Memory容量2GBである。実験で使用した各閾値を、表1に示す。実験に使用した映像は、一般個人が撮影した映像と、比較のために、テレビ放送におけるドラマ映像と監視カメラのように全くカメラに動きがなく、画面内のオブジェクトに全く動きのない映像の3つを用意した。画像サイズは 352×240 、再生時間は60秒である。実験結果を表2に示す。計算フレーム数は、動きベクトルを算出したフレームの数を示す。

表2より、映像の内容によって、処理速度と計算フレーム数が大きく変化することがわかる。監視カメラ映像のように、明らかに全てのカメラワークがFIXである映像では、 INT_1 フレームに1回しか動きベクトルを算出していないためであり、処理時間もリアルタイム再生時間の25%程度とかなり高速である。逆に、

表1 実験で使用した各閾値

INT_1	15
INT_2	5
Th_{fix}	20%
l_{search}	36
$l(INT_1 \text{ 使用時})$	12
$l(INT_2 \text{ 使用時})$	8
Th_{good}	3
Th_{merge}	9
Th_{bad}	9

表2 実験結果

映像種類	処理時間	計算フレーム数
一般個人映像	77 秒	2286 回
ドラマ映像	41 秒	1040 回
監視カメラ映像	15 秒	123 回

FIXの比較的少ない一般個人映像では、計算フレーム数も多く、リアルタイム再生速度を超えてしまっている。ドラマ映像は、その中間の値となっていることがわかる。これは、一般個人の映像よりもFIXである部分が多いことが処理時間に影響していると考えられる。

7. まとめと今後の課題

本稿では、映像シーケンスからのカメラワーク情報抽出とその情報を利用したセグメント分割法とカメラワーク安定性評価決定法について、アルゴリズム検討を行った。処理時間に関する簡易実験を行ったところ、一般個人撮影の映像に対してリアルタイム再生時間の約1.3倍程度の時間で処理できることが確認することができたが、システムの目的を考慮し、より高速化を検討する必要がある。また、抽出したカメラワークの正確性や算出したカメラワーク安定度評価値の妥当性に関する検討も、今後の課題である。

文献

- [1] 安田浩, "コンテンツ流通に向けた取り組みと課題," 2001年映像情報メディア学会年次大会, S2-1, pp.405-410, 2001
- [2] 金田瑞規, 高木真一, 小館亮之, 富永英義, "映像編集支援システムのためのインデキシングパラメータ抽出手法の検討," 情報処理 Vol.2001, No.121, 2001-AVM-35, pp.13-18, 2001
- [3] Changming Sun, "Fast Optical Flow Using Cross Correlation and Shortest-Path Techniques," Digital Image Computing: Techniques and Applications, pp143-148, 2001