

放送済み番組の再契約を目的としたオーサリングの 為の映像同定技術に関する検討

横山 和正[†] 高木 真一^{††} 小舘 亮之^{†††} 富永 英義^{††,†††}

[†] 早稲田大学大学院 理工学研究科

〒169-8555 東京都新宿区大久保 3-4-1

^{††} 早稲田大学 大学院 国際情報通信研究科

〒169-0051 東京都新宿区西早稲田 1-3-10

^{†††} 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

E-mail: kyokoyama@tcn.comm.waseda.ac.jp

あらまし ブロードバンドネットワークの普及とともに、過去の放送済み番組を再配信等に利用するケースが増加している。しかし、再配信のためには著作権等の権利保有者と再契約を行う必要があり、番組内容の資料作成のための作業が大きな負担となっている。そこで我々は、放送済み番組のレンタル映像等の外部の映像データを利用している部分を検出する映像同定技術に関する検討を行っている。本稿では、このような同定技術のための映像特徴量に関する基礎実験を行ったので報告する。

キーワード 映像コンテンツ, 映像同定, 相関係数, 外部映像, MPEG

A Basic Study on Video Identification Method for Authoring of Used TV programs

Kazumasa YOKOYAMA[†], Shin'ichi TAKAGI^{††}, Akihisa KODATE^{†††}, and Hideyoshi
TOMINAGA^{††,†††}

[†] Department of Electronics, Information and Communication Engineering, Waseda University
Okubo 3-4-1, Shinjuku-ku, Tokyo, 169-8555 Japan

^{††} Graduate School of Global Information and Telecommunication Studies, Waseda University
Nishiwaseda 1-3-10, Shinjuku-ku, Tokyo, 169-0051 Japan

^{†††} Global Information and Telecommunication Institute, Waseda University
Nishiwaseda 1-3-10, Shinjuku-ku, Tokyo, 169-0051 Japan

E-mail: kyokoyama@tcn.comm.waseda.ac.jp

Abstract Today, with the spread of broad band network, re-distribution of used TV programs is increasing. However, to make a re-contract with rights holders for re-use TV programs is necessary, and making a detailed data of it is very troublesome work. Then we are studying a method of video identification, to make it possible to detect the same video scene between used TV programs and the other video data. In this paper, we report a result of basic study of significant information for such a video identification.

Key words Video content, Video identification, Correlation coefficient, MPEG

1. はじめに

CSデジタル放送、BSデジタル放送や、本年末の地上波デジタル放送の本放送の開始に代表される様に、現在、放送のデジタル化が進みつつある。この、放送のデジタル化の特徴の一つとして、多チャンネル化をあげる事ができる。また、通信網の広帯域化に伴ない、既存の放送波以外の映像配信可能なインフラとして、ブロードバンドインターネットが期待されている。この様に、近年、映像の伝達手段が増加しつつある。

映像の伝達手段が増加すれば、伝達する中身である番組や、映像コンテンツの必要数も増加する事が考えられる。しかし、放送で使用されている番組の作成には、多大なコストがかかるのが現状であり、映像の伝達手段の増加に合わせて新しく番組を作成しては、コストの面で負担が大きくなってしまう。

そこで、新しく番組を作成するだけではなく、過去に放送済みの番組を、再び放送や映像配信に使用する需要が高まっている。しかし、一般的に、放送番組はこの様な再利用、特にブロードバンドインターネットでの配信などにおける再利用を前提としていないので、放送済みの番組を再び使用する為には、その番組に関わる著作権者と再契約を行う必要がある。再契約を行う為には、番組中のどの映像が借用して使用されたものか、どの様な出演者が登場しているのか、どの様な楽曲が番組中において使用されているのか、等の権利に関する情報を明確にする必要がある。

それに対して、放送済み番組に関する資料として残されているものは、放送済み番組のテープと番組作成時の契約書程度しか存在せず、再利用する際には、目視によるテープの内容確認を行う必要がある。この場合、資料の作成自体に大きな手間を必要とし、放送済み番組を再利用する際の大きな問題となっている。

その為、放送済みの番組を有効利用する為に、番組内容に関する資料作成のオーサリングツールの必要性が高まりつつある。

そこで、本研究では、この様なオーサリングを可能にする技術の一つとして、映像同定技術を用いて、放送済み番組中の、レンタル映像等の外部の映像データを利用して映像部分の検出の検討を行う。

2. 映像の同定技術

2.1 一般的な映像の同定技術

映像の同定に関する研究としては、多くのものがある事ができる[1],[2],[3],[4],[5],[6]。しかし、その研究の多くは、放送されている映像ストリームや、アーカイブ中の映像を検索、検出する事が目的である。その為、図1の様に、検索、検出を行いたい参照映像を所有している事が前提で、その参照映像と

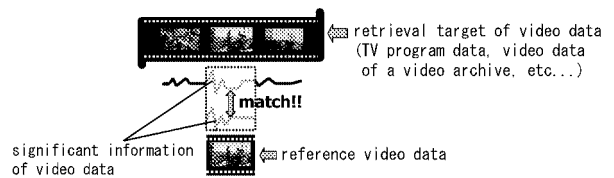


図1 一般的な映像の同定技術

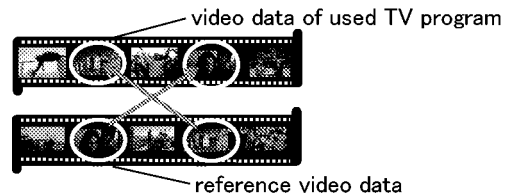


図2 本研究で検討する映像の同定技術

同じ部分を検索対象とする映像から探すものであり、検索対象映像中に、どの様な映像が存在するか、という事は既知である。

2.2 本研究における映像の同定技術

本研究において検討する映像の同定技術について述べる。

まず、本研究においては、

- 放送済み番組の映像データ
- 放送済み番組が映像を一部参照した、もとの映像データ

の2種類の映像データが手元に存在する事を前提とする。すなわち、前述した番組作成時の契約書などから、放送済みの番組は、後者の映像データより映像を一部参照している事は既知であるものとする。

但し、前節の映像の同定とは異なり、この映像データの双方に、どの様な映像が共通して存在するか、という点は、未知である。

その為、この様に放送済み番組の映像データと、参照もとの映像データが手元に存在する中、

放送済み番組中のどこにどれだけ、外部の映像のどの部分が使用されているのかわからない

という点が、問題点として発生する事となる。

そこで、本研究では図2の様に、放送済み番組のどこにどれだけ、参照もとの映像のどの部分が存在するかを調べる、すなわち、複数の映像データ間の共通部分を検出する、という映像の同定技術について検討を行う。

3. 映像の特徴量抽出手法について

3.1 特徴量を抽出する単位

本研究と類似した映像の同定の実現を試みているものとして、文献[4],[5],[6]が挙げられる。これらの関連研究においては、映像から特徴量を抽出する単

位をフレーム単位で行い、その比較もフレーム単位、たとえ複数フレーム単位であっても、高々30フレーム単位であった。

フレーム単位で比較を行えば、正確に映像データ間の一致箇所を検出することは可能である。しかし、フレーム単位での特徴量の比較は、正確な検出を実現する反面、その計算量は非常に大きなものとなってしまう。

また、本研究においては、番組内容に関する資料作成のオーサリングを可能にする技術を目指している。この場合、全ての作業が自動化され、全く人手が介されなくなる事は考えがたく、最終的には人手が入る事が予想され、その、人手で行われる作業を支援する事が要求されると考えられる。その為、映像間の一致部分の検出も、フレーム単位で行う必要は無く、大体の一致部分を検出するだけで十分支援が行える事と予想できる。

そこで、本研究では、フレーム単位で特徴量を抽出し、比較を行うのではなく、まず、映像データをショット単位で分割し、そして分割したショットからそのショットを代表する特徴量を抽出して比較を行う事で、比較を行う映像中のフレーム一つ一つから特徴量を抽出せずに、映像間の大体の一致部分の検出を行う事を目指す。

3.2 ショットの特徴

映像データ中のショットの映像は、

- 時間方向の色の変化が少ない
- 時間長の短いものを含む

という特徴を持っている。

前者は、ショットの切れ目であるカット点やシーンの切れ目であるシーンチェンジは、もともと映像データ中において、色が大きく変化する点であった為である。

また、後者は、映像データをショット単位で分割した場合、映像データ中に含まれるショットの長さは、長いものから短いものまで様々なものが含まれていると考える事ができる為である。

3.3 抽出する特徴量に対する考察

ショットの様な、映像中の複数のフレーム群を代表する特徴量として、映像中のフレーム毎に色の平均値を算出し、それを時系列に並べ、その時間方向の変化をフレーム群の特徴とするものを挙げる事ができる[1][2]。

しかし、ショットは、前節で述べた特徴を有する為、この様な時間方向の変化を特徴とする特徴量は、特徴となる部分が少なくなり、映像間の特徴量の違いも減少する。その為、ショットの比較には、この様な時間方向の特徴を抽出する特徴量は適さず、空間方向の特徴を抽出する特徴量が適していると考えられる事ができる。

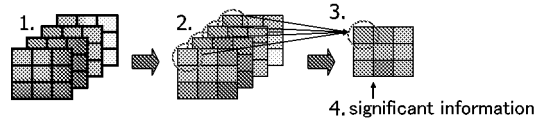


図3 提案する特徴量の抽出手法

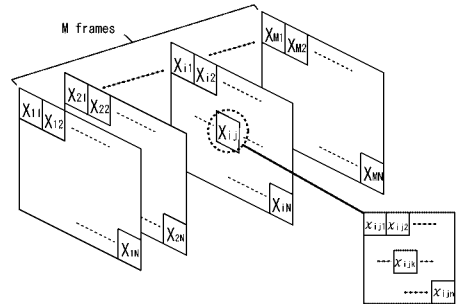


図4 特徴量の抽出の為のパラメータ

3.4 使用する特徴量の抽出手法

前節の考察を踏まえ、空間方向の特徴を抽出する特徴量として、文献[7]において提案したものを使用する。この特徴量は、ショットのフレームを複数のブロックに分割、各ブロック毎の時間方向、空間方向の色の平均値をとったものの集合である。

具体的な特徴量の抽出手法を以下に示す。

- (1) フレームを複数のブロックに分割する (図3.1)
- (2) ブロック毎に輝度、色差の平均値を算出 (図3.2)
- (3) フレーム間の各ブロックの輝度、色差の平均値を算出する (図3.3)
- (4) 各ブロックの輝度、色差の平均値の集合を特徴量とする (図3.4)

また、図4の様な、 M フレームのショットを、ブロックの分割数を N として抽出した特徴量 A は、次式で表される事になる。但し、 n は分割したブロック中の画素数であり、 x_{ijk} ($k=1, 2, \dots, n$) はショット中の i フレーム目の、分割したブロック中の j 番目のブロックの各画素値である。

$$A = \{ \bar{X} | \bar{X} = \frac{\sum_{i=1}^M X_{ij}}{M}, (i=1, 2, \dots, N) \} \quad (1)$$

$$X_{ij} = \bar{x}_{ij} = \frac{\sum_{k=1}^n x_{ijk}}{n} \quad (2)$$

4. 特徴量の比較手法の検討

4.1 従来手法

4.1.1 平均二乗誤差法を使用した比較手法

前節の手法で映像から抽出した特徴量は、数値系列と見なす事ができる。この様に、映像から抽出し

た特徴量を数値系列と見なして比較する手法の一つとして、平均二乗誤差法を挙げる事ができる [4], [7]. そこで、従来手法として、平均二乗誤差法を特徴量の比較に用いた場合、本研究が目標とする映像の同定を実現できる可能性があるか検証する。

4.1.2 平均二乗誤差の算出方法

平均二乗誤差の値の算出方法は、次の様になる。

二つのショット A_i, B_i から抽出した特徴量 A, B は、輝度、色差に対して、各々図 5 の様に、格子状に各ブロックの値 $a_1, a_2, \dots, a_N, b_1, b_2, \dots, b_N$ が並んでいると考えることができる。この場合、特徴量 A, B の平均二乗誤差 ME は、

$$ME = \sum_{i=1}^N \frac{((b_i - a_i) - ave)^2}{N} \quad (3)$$

$$ave = \sum_{i=1}^N \frac{|b_i - a_i|}{N} \quad (4)$$

となる。但し、ここで N は、フレームの分割数である。

4.1.3 平均二乗誤差法についての考察

平均二乗誤差法は、二つの特徴量を比較する際によく使用される手法であり、特徴量の数値系列同士の値がどれだけ一致しているか、という度合いを示す。その為、図 6.1 の様に、放送済み番組中の参照した部分であるショットが、参照したもとの映像データのショットと一致する場合の比較に用いるのには適している。しかし、図 6.2 の様に、参照した映像が、参照もとの映像のショットの一部となっている場合は、その両者の特徴量は、完全に一致しない為、平均二乗誤差の値を算出して比較を行う事は適さないと考えられる。

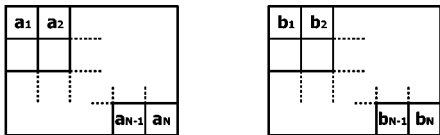


図 5 ショットから抽出した特徴量

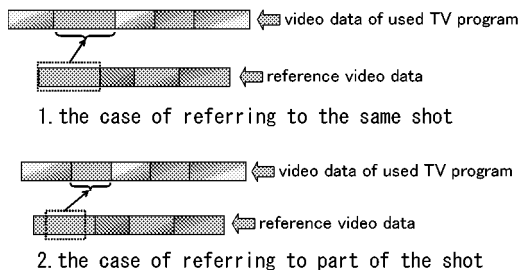


図 6 映像の参照方法

表 1 平均二乗誤差の算出値 (その 1)

	時間長 1/2		
	輝度 Y	色差 Cb	色差 Cr
Fix のショット	156.1	3.296	9.636
動きのあるショット	1281	20.95	22.95

4.1.4 平均二乗誤差法による実験

前節の考察結果を確認する為に、平均二乗誤差法を用いて図 6.2 の様な、ショットの一部を切り出したショットの特徴量と、切り出しもとのショットの特徴量の比較を行った。

用意した映像は、

- カメラワークが全て Fix であるショット
- Pan, Zoom 等、動きのあるカメラワークを含むショット

の 2 種類を用意し、その各々に対して三種類のショットを用意、各々ショットとその一部の映像の特徴量の平均誤差を算出した。但し、特徴量を抽出する際のフレームの分割数 $N=330$ (縦 15 × 横 22) とし、ショットの一部の映像には、もとのショットの時間長が 2 分の 1, 4 分の 1 の映像を用意した。また、カメラワークの判定には、文献 [8] の手法を用い、実験に使用した映像は映画「Perfect Storm」のショットを使用した。

実験結果を表 1, 表 2 に示す。表 1, 表 2 の各値は、各々の映像データで算出した平均二乗誤差の値の平均値である。

また、平均誤差の値の指標として、共通するフレームが存在しないショットの特徴量同士を比較した場合の平均二乗誤差の値を表 3 に示す。

4.1.5 結果に対する考察

表 1, 表 2, 表 3 を比較すると、カメラワークを持つショットの場合は、切り出した時間長が短くなると、平均二乗誤差の値が共通フレームが存在しないショット同士を比較した場合の平均誤差の値に近くなり、映像の同定を行う事が難しくなっている事がわかる。

また、ショット内のカメラワークが Fix の場合についても、カメラワークがある場合程度ではないが、平均誤差の値は、時間長が短くなるにつれて高くなっている事がわかり、特に、時間長が 4 分の 1 の時は、共通フレームが存在しないショット同士を比較した場合よりも平均二乗誤差の値は小さいものの、明確な差があるとは言えない事がわかる。

以上の結果より、特徴量の比較に、平均二乗誤差法を使用した場合、時間長を短くすると、特にショット中にカメラワークを含む場合は同定を行なう事ができず、ショット中のカメラワークが Fix の場合も、明確な差を示して映像の同定を行う事ができない事がわかる。

表 2 平均二乗誤差の算出値 (その 2)

	時間長 1/4		
	輝度 Y	色差 Cb	色差 Cr
Fix のショット	49.9	9.301	25.29
動きのあるショット	1818	36.93	27.37

表 3 平均二乗誤差の算出値 (その 3)

共通フレームが存在しないショット同士	輝度 Y	色差 Cb	色差 Cr
	164	70.89	81.77

4.2 相関係数を使用した比較手法

4.2.1 比較する特徴量の散布図

前節において、ショットとその一部を切り出した映像の特徴量の間には、ある程度の類似性はあるが、平均二乗誤差法ではそれを明確に示す事ができなかった。そこで、双方の類似性を確認する為に、ショットとその一部を切り出した映像の特徴量の値の散布図を調べる。

図 7 にショット中のカメラワークが全て Fix の場合を、図 8 にショット中のカメラワークに動きがあるものを含む場合を、図 9 に共通ショットが存在しない場合の輝度 Y についての特徴量の散布図を示す。

一般的に散布図において、ある二つの数値系列の分布が、縦軸と横軸の間に直線的に分布が集まっている場合、その二つの系列は相関性を有するとされている [9]。

図 7 より、ショット中のカメラワークが全て Fix である場合は、比較に用いた二つの特徴量は、相関性を有している事がわかる。また、図 8 より、ショット中にカメラワークを含む場合は、縦軸と横軸の間にある程度分布が集まってはいるが、離散している部分もあり、図を見ただけでは明確にはわからない。図 9 より、共通するフレームが無い場合は、その分布は離散しており、二つの特徴量の間に相関性は見られない事が示されている。

4.2.2 相関係数の算出方法

散布図から読み取った相関性を数値的に示す為に、各々の場合の相関係数を算出する。ここで、相関係数 r は、次式で示されるものである。

$$r = \frac{\frac{1}{N} \sum_{i=1}^N (a_i - \bar{a})(b_i - \bar{b})}{\sqrt{\frac{1}{N} \sum_{i=1}^N (a_i - \bar{a})^2} \sqrt{\frac{1}{N} \sum_{i=1}^N (b_i - \bar{b})^2}} \quad (5)$$

また、相関係数 r の取る値の範囲は、-1 から 1 までであり、相関係数が 1 ならば、まったく同じ数値

表 4 相関係数 r と相関性の関係

相関係数 r の値	相関性
0~0.2	ほとんど相関が無い
0.2~0.4	やや相関がある
0.4~0.7	かなり相関がある
0.7~1.0	強い相関がある

の系列である事を示す。一般的には、相関係数 r と相関性の関係は、表 4 の様になるとされている [9]。

4.2.3 相関係数による実験

平均二乗誤差を用いた比較実験と同じ条件で、相関係数を用いたショットの特徴量の比較の実験を行った。

表 5、表 6、表 7 に相関係数の平均値の算出結果を示す。また、輝度 Y、色差 Cb、Cr を RGB 色情報に変換した場合の相関係数の平均値の算出結果を表 8、表 9、表 10 に示す。

表 5 輝度、色差の相関係数の算出値 (その 1)

	時間長 1/2		
	輝度 Y	色差 Cb	色差 Cr
Fix のショット	0.9715	0.9609	0.9381
動きのあるショット	0.8443	0.5268	0.5015

表 6 輝度、色差の相関係数の算出値 (その 2)

	時間長 1/4		
	輝度 Y	色差 Cb	色差 Cr
Fix のショット	0.9224	0.86825	0.8283
動きのあるショット	0.8255	0.5655	0.5256

表 7 輝度、色差の相関係数の算出値 (その 3)

共通フレームが存在しないショット同士	輝度 Y	色差 Cb	色差 Cr
	0.7532	-0.1507	-0.1008

表 8 RGB の相関係数の算出値 (その 1)

	時間長 1/2		
	R	G	B
Fix のショット	0.9667	0.9727	0.9725
動きのあるショット	0.8400	0.8430	0.8342

表 9 RGB の相関係数の算出値 (その 2)

	時間長 1/4		
	R	G	B
Fix のショット	0.9109	0.9249	0.9284
動きのあるショット	0.8235	0.8218	0.8104

表 10 RGB の相関係数の算出値 (その 3)

共通フレームが存在しないショット同士	R	G	B
	0.6166	0.5817	0.5644

4.2.4 結果に対する考察

表 5、表 6 より、ショット中のカメラワークがすべて Fix の場合は、輝度 Y、色差 Cb、Cr とともにほぼ 0.9 以上、少なくとも 0.8 以上の値をとっており、相関性が高い事が示されている。

また、ショット中のカメラワークに動きがあるものを含む場合は、共通するショットが無い場合よりは高い相関係数の値をとっているものの、相関係数

そのものは比較的低い値を、特に、色差が低い値をとっている。

さらに表7より、共通フレームが存在しないショット同士の相関値は、特に色差は低い値をとっている事がわかる。

以上の事より、輝度 Y 、色差 Cb 、 Cr の相関係数の値を比べる場合、ショット中のカメラワークが全て Fix の場合は、ショットとその一部を切り出した映像同士の同定を行える可能性がある事が示せた。しかし、カメラワークに動きがあるものを含む場合は色差の相関係数が低くなっており、また、表5、表6、表7の各々の輝度 Y の値は全て、0.7以上であり、表4の、強い相関がある、という分類に入ってしまったという結果となっている。

そこで、 RGB の相関係数を算出した、表8、表9、表10を見ると、 RGB の相関係数は、カメラワークが全て Fix のショット、カメラワークに動きのあるショットのどの場合も相関係数の値が0.8以上であり、さらに、共通フレームが存在しないショット同士の場合は、 R 、 G 、 B いずれの値も0.7未満であり、表4の分類においても異なる分類に位置している為、映像の同定を行える可能性がある事を示せた。

5. ま と め

放送済み番組の再契約の資料作りを目的としたオーサリングを可能にする技術の一つとして、映像同定技術を用いて、放送済み番組中のレンタル映像等の、外部の映像データを利用している映像部分の検出の検討を行った。

映像の同定技術として、複数の映像データ間の共通する映像部分を検出する手法について検討し、映像をショット毎に分割し、そのショットを代表する特徴量を抽出して比較を行う手法を提案した。

そして、放送済み番組中で参照されている映像が、もとの映像中のショットの一部を切り取ったものであっても映像の同定が行える特徴量の比較手法として、 RGB 色情報の相関係数を使用する手法を提案し、その有効性を確認した。

今後の予定として、以下のものを検討している。

- カメラワークに動きを含んでいるショットの場合の、より明確な結果が得られる比較手法の検討
- 計算速度も考慮した映像の同定手法の提案
- 上記手法の実験及び評価

文 献

- [1] 高橋克直, "有意情報抽出による画像情報へのアクセス手法の研究," 平成12年度早稲田大学博士論文, (2001)
- [2] 高橋秀和, 児玉明, 他, "局所領域特徴量を用いた動画画像検索方式," 信学技報 IE2001-196, pp.37-42, (2002)
- [3] 黒住隆行, 柏野邦夫, 村瀬洋, "携帯カメラで収録した映像をキーとする一致映像探索," 信学技法 PRMU2002-130 pp.7-11, (2002)

- [4] 粕谷英司, 他, "MPEG-7 による映像制作支援素材アーカイブシステム," 情処研報 AVML-40, pp.49-54, (2003)
- [5] 片山紀生, 佐藤真一, "映像断片照合を用いた大規模ニュース番組映像アーカイブの構造化," 信学技報 PRMU2002-23, pp.21-26 (1996)
- [6] 山岸史典, 佐藤真一, 浜田喬, "大規模放送映像アーカイブにおける映像断片照合の提案と高速化," 信学技報 PRMU2002-166, pp.121-126, (2002)
- [7] 横山和正, 高木真一, 小館亮之, 富永英義, "映像コンテンツ同定のための特徴量抽出手法," 情処研報 AVML-36, pp.7-12, (2002)
- [8] 土橋健太郎, 小館亮之, 富永英義, "手ぶれを考慮した MPEG2 からのカメラワーク検出の検討," 信学技報 IE2000-100, pp.1-6, (2001)
- [9] 石村貞夫, "統計解析のはなし" 東京図書, (1989)

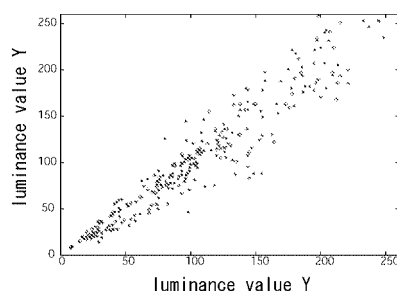


図7 カメラワークが全て Fix の場合の散布図

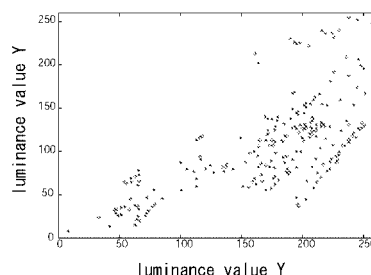


図8 カメラワークに動きがある場合の散布図

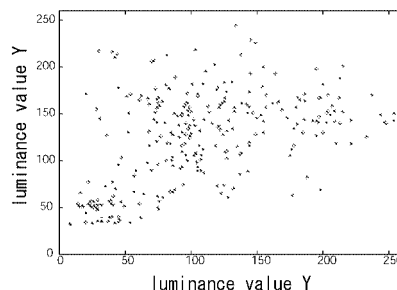


図9 共通ショットが存在しない場合の散布図