

部分空間の正準角による 3 次元パターンのマッチング法

前田 賢一† 山口 修† 福井 和広‡

† (株) 東芝 研究開発センター

〒 212-8582 川崎市幸区小向東芝町 1

E-mail: {ken.maeda, osamu.yamaguchi}@toshiba.co.jp

‡ 筑波大学 システム情報工学研究科

〒 305-8573 つくば市天王台 1-1-1

E-mail: kfukui@cs.tsukuba.ac.jp

3次元の物体認識のためには、3次元のパターン間の類似性を定義する必要がある。複数方向からカメラで撮影された3次元のパターンを対象として、3次元パターン同士をマッチングする方法について考察する。マッチングの方式として、3次元パターンをカメラで撮影した画像の集合を部分空間で近似し、部分空間の間の正準角を利用したマッチングを提案する。また、その効果を実験によって確認する。

A 3-Dimensional Pattern Matching Using Canonical Angles

Ken-ichi Maeda†, Osamu Yamaguchi†, and Kazuhiro Fukui‡

† Corporate Research and Development Center, TOSHIBA Corporation

1, Komukai-toshiba-cho, Saiwai-ku, Kawasaki, 212-8582, Japan

E-mail: {ken.maeda, osamu.yamaguchi}@toshiba.co.jp

‡ Graduate School of Systems and Information Engineering, University of Tsukuba,

1-1-1, Ten-noudai, Tsukuba, 305-8573 Japan

E-mail: kfukui@cs.tsukuba.ac.jp

3-dimensional pattern recognition requires the definition of a similarity measure between 3-dimensional patterns. We discuss how to match 3-dimensional patterns, which are represented by a set of images taken from multiple directions and approximately represented by subspaces. The proposed method is to calculate the canonical angles, in particular the third smallest angle between two subspaces. We demonstrate the viability of the proposed method by performing a pilot study of face recognition.

1 まえがき

近年、パターン認識で扱われる問題は、2次元から3次元へと範囲をひろげつつある。具体的な目的のひとつとしては、写真と実物とを見分けるといような応用を想定することができる。必然的に3次元のパターンを認識する必要が生じ、その具体的な方式が必要とされている。

パターン認識のためには、一般に、パターン間の類似性を定義することが必要であり、その類似性の程度によって認識結果を決定するというプロセスがとられる。

したがって、3次元の物体を対象とする場合には、まず、3次元のパターン間の類似性を定義する必要がある。

3次元パターンを表現する方法はいろいろ考えられ、たとえば、3次元空間のボクセルに物体の有無を割り当て、表面にはテクスチャーを貼り付けるというように、3次元のモデルを使うのも、ひとつの方法である。しかし、物体を外側から見るとい観測方法を前提とするならば、物体内部の情報は、それほど重要ではなく、表面の情報のみが意味を持つことになる。また、表面の情報に限定したとしても、こうした3次元のモデ

ルを作成するのは、一般に計算コストのかかる処理である。さらに、モデルができたとして、モデル同士の間、いかに類似性を定義すればよいかという問題も残されている。

一方、外側から見ることを前提とした表現法のひとつとして、パラメトリック固有空間法 [1] が知られている。それは、3次元の対象物を複数の方向から観察された画像の集合で表現するものである。この集合は、固有空間上で多様体上に分布するため、一方向から観測して得られた画像との距離を、多様体と点との距離として定義し、それが一方向から観測された画像と3次元物体とのマッチングの尺度となる。

しかし、この方式では、あらかじめ登録しておく対象は3次元を反映しているものの、1回の認識対象はカメラで一瞬を撮影したもので、すなわち2次元であった。このため、前述の応用例で、写真と実物とを見分けることを想定した場合、入力固有空間内の一点で表現されるものであるとすると、たまたまある方向から撮影された写真と、同じ方向から観測された実物とを見分けることはできないことになり、別の工夫が必要となる。

ここで提案するのは、こうした問題を解決するための3次元パターンのマッチング方法である。すなわち、リアルタイムに近い速度の処理が可能であり、具体的な応用として、3次元の対象物と写真との区別をつけることが可能であるような方式である。

基本的なアイデアは、パラメトリック固有空間法と同様に、複数の方向から観測された画像の集合を3次元物体の表現の基本とし、それを部分空間で近似して、部分空間の間の正準角（特に3番目に小さい正準角）を利用するというものである。これは、2次元のパターン認識で使われた相互部分空間法 [2] を3次元に拡張したものである [3]。

2 問題へのアプローチ

2.1 基本方針（部分空間の利用）

3次元パターンのマッチング方法は、従来の2次元パターンのマッチング方法の自然な拡張であることが望ましい。たとえば、部分空間法は、よく知られた2次

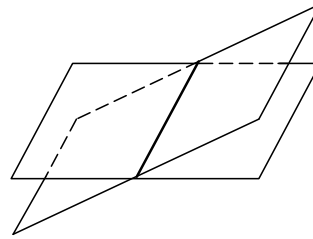


図 1: 3次元中の2平面

元パターンのマッチング方法である [4]。部分空間法では、高次元空間のパターンの分布を低次元の部分空間で近似することを前提とする。

これを拡張した相互部分空間法 [2] では、パターンの分布を部分空間で近似して、辞書、入力ともに変動するパターン同士の類似度を、2つの部分空間の間の最小の角度で定義する方式が提案された。これは顔認識にも応用された [5]。さらに環境条件の変動を吸収する工夫も加えられ、最も近いところで測るという観点に立つものとしては、十分実用的であることが示されていた [6]。

しかし、これらの方式で使われていたのは、2つの部分空間の間の最小の角度であった。たとえば、3次元空間の中に異なる2つの平面を置いたとすると、必ず一直線を共有して図1のようになる。部分空間として異なっているわけであるが、従来のように最小の角度を測ると0度となり、完全にマッチングがとれたという評価になる。

この現象は、2個の3次元物体がある方向から見て、たまたま同じように見えた場合に同じ物体とみなされることに対応しており、このままでは、やはり、3次元物体とその写真とを区別できない例が存在することになる。

2.2 複数の正準角の利用

部分空間の間の角（正準角）とは、互いに直交する方向で測った部分空間が交わる角度である [7]。正準角を求めるには、射影行列の積 PQP , QPQ などの固有値を計算し、これが正準角の余弦の2乗であることを使う。この行列の次元数が高いため、実際には、等価な問

題に変換し，

$$X = (x_{ij}) \quad (1)$$

ただし，

$$x_{ij} = \sum_{m=1}^M (\psi_i, \phi_m)(\phi_m, \psi_j) \quad (2)$$

または

$$x_{ij} = \sum_{n=1}^N (\phi_i, \psi_n)(\psi_n, \phi_j) \quad (3)$$

なる行列 X の固有値を計算する [2]．ただし， M 次元の辞書部分空間の基底ベクトルを $\{\phi_m\}_{m=1}^M$ ， N 次元の入力部分空間の基底ベクトルを $\{\psi_n\}_{n=1}^N$ とする．

正準角を $\{\theta_k\}_{k=1}^K$ とすると，

$$Xz = \mu_k z \quad (4)$$

$$\cos^2 \theta_k = \mu_k \quad (5)$$

として求めることになる．文献 [2] では，最大固有値（すなわち最小の正準角）だけを求めた．

一方，文献 [8] では，2つの部分空間への射影を，それぞれ P, Q としたとき，

$$PQPy = \mu_{\min} Py \quad (6)$$

の最小固有値（すなわち最大の正準角）を求めた．ただし，

$$\|Py\| = 1 \quad (7)$$

とする．この式 (7) の条件がないと，最小の固有値は 0 であることが多く，かならずしも部分空間の間の最大の正準角を反映しない．

ここで考察対象として仮定するのは，固定されたカメラで 3 次元物体を撮影し，その物体の向きの変化が微小であるということである．具体的なターゲットとして想定されるのは，たとえば顔認識である．顔を顔と特定するために顔の部品として，両目と両方の鼻孔を使うことを想定すると [5][9]，顔向きの変化は，これらの部品がカメラから見える範囲に限定されることを仮定することになる．また，抽出された顔領域の内，認識に使われる領域は，大きさをそろえるという正規化を施すものと仮定する．

そうすると，得られた認識に使われる領域の変化は微小なものになる（図 2 参照）．変化が微小であるとす



図 2: 正規化後の認識領域

ると，ある方向（たとえば横向き）の回転によって変化する認識領域パターンの集合が 2 次元の部分空間上に分布していると近似することができる．また，別の方向（たとえば縦向き）の回転によって変化するパターンの集合は別の 2 次元の部分空間上に分布していると考えられる．

実際には，顔を対象とした場合には，表情の変化などが原因で，回転も含めて N 次元の部分空間で近似される分布になっている．しかし，表情の変化は同じ表情をすることは不可能に近いので毎回違ったものになるのに対して，回転は再現が容易であると考えられる．あるいは，回転方向がずれたとしても，微小な回転であれば，横と縦の回転の合成で近似できると考えることができる．そうすると，任意の方向に回転させながら収集した 2 組の顔画像シーケンスに対して，部分空間で近似したとすると，2 次元分はほぼ一致していると考えても良い．さらに，顔の回転が 1 方向でない場合に関しては，3 次元分がほぼ一致していると考えても良い．

この考察から得られるものは，2つの部分空間のなす正準角の内，3個はほぼ 0 であるということである（ただし，それは表情の変化がない場合であり，表情の変化がある場合には，それに応じて角度が大きくなることになる）．したがって，3個の正準角を利用するという方式を提案することができる．

3次元での回転は，もうひとつあり，カメラの光軸の回りの回転であるが，これは，考慮する問題として 3 次元物体とその写真を区別するということ考えた場合，回転したものが同じになるため，考慮しないものとする．この様子を図 3 に示す．すなわち，カメラが顔のほ

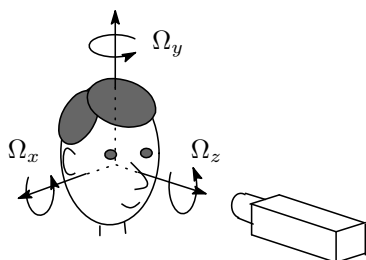


図 3: 3次元物体の回転

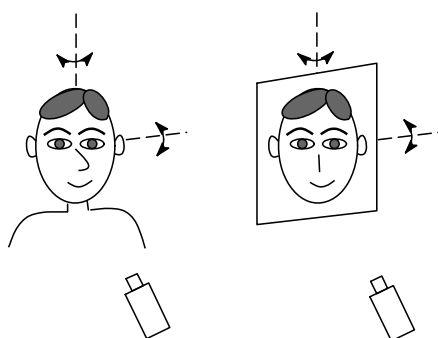


図 4: 実物と写真の区別問題

ば正面にあるとすると、考慮する回転は、 Ω_x 、 Ω_y だけであり、 Ω_z は考慮しない。

さらに、正準角の内、小さい方から3個を利用するという枠組みの中で、いくつかの方法を考えることができ、たとえば平均を計算する方法 [10] も考えられる。平均を取ると、大きな値が支配的になるため、ここでは、文献 [8] にある最大の角度という考え方を再検討してみる。これは、ノルムの定義としての最大値ノルムの考え方である。すなわち、3個の中で最小のもの（角度でいうと最大のものに相当）を利用するというのが、妥当であると思われる。

2.3 3次元物体と写真の差

思考実験として、3次元物体と、その物体を1方向から撮影した写真を区別するという課題を考えてみる（図4参照）。

前述の通り、正面から撮影した1枚の画像のみからでは、この区別をつけることができない場合がある。複数の方向から撮影した画像を利用するとすると、実際の物体では、撮影する角度の変化によって、見えてい



図 5: 顔の例



図 6: 顔写真の例

た部分が隠れたり、見えていなかった部分が現われたりする。しかし、写真では同じ物が affine 変換されて見えるだけある（カメラモデルとして平行投影を仮定）。この差が、部分空間の間の正準角の内、最小の角度以外に反映する。ここでは便宜的に対象は静止していて、カメラが動くように説明したが、実際的にはカメラが静止していて、物体が動く場合もある。このような場合には、実際の物体に対しては光の当たり方が変化することによる見え方の変化も重畳される。写真では全体の明るさが変化するだけである（照明として平行光線を仮定）。

3 基礎実験

前述の思考実験に対して、第3正準角の効果を検証するため、第1正準角との比較を行なった。P0からP10までの11名の顔の中で、P0の顔を登録し、全員の実際の顔（図5）と、写真（6）をインプットした場合の値の比較である。表1に最小正準角（最大固有値 μ_1 ）を使った場合の類似度、表2に3番目に小さい正準角（3番目に大きい固有値 μ_3 ）を使った場合の類似度を示す。

最大固有値（ μ_1 ）を比較してみると、3次元物体で0.989であるのに対して、写真でも0.977という高い値である。これを分離する閾値を設定するのが困難であることがわかる。他方、第3固有値（ μ_3 ）を比較してみると、3次元物体で0.937であるものが、写真では0.204となっており、十分大きな差がある。両方とも、本人（P0）と他人を区別するという点では、十分大きな差が認められる。

表 1: 最小正準角 (最大固有値 μ_1) による認識結果

Person	顔入力	写真入力
P0	0.989	0.977
P1	0.702	0.591
P2	0.707	0.619
P3	0.786	0.741
P4	0.701	0.665
P5	0.643	0.626
P6	0.730	0.612
P7	0.554	0.678
P8	0.750	0.732
P9	0.716	0.600
P10	0.772	0.648

表 2: 第 3 正準角 (第 3 固有値 μ_3) による認識結果

Person	顔入力	写真入力
P0	0.937	0.204
P1	0.256	0.165
P2	0.520	0.237
P3	0.488	0.123
P4	0.457	0.075
P5	0.459	0.124
P6	0.227	0.055
P7	0.334	0.238
P8	0.557	0.246
P9	0.545	0.154
P10	0.435	0.075

4 拡大実験

基礎実験では、一人の登録辞書に関して、11 名の顔と写真の入力に対して、最大固有値と第 3 固有値の値を比較した。ここでは、基礎実験とは別の 11 名を登録し、最大固有値を用いた場合と第 3 固有値を用いた場合とで効果を確認する。

表 3 に、それぞれ、写真と顔の入力に対する類似度の分布を表示する。ここで注目するのは、顔および写真の入力に対する本人の類似度の状況である。最大固有値を用いた場合には、顔入力に対する本人の類似度が最も小さいケースは 0.970 (m011real) であるのに対して、写真入力に対する本人の類似度が最も大きなケースは 0.965 (m006photo) となっている。一方、第 3 固有値を用いた場合には、それぞれ、0.905 (m011real) と 0.616 (m004photo) である。この実験でも、最大固有値では実物と写真を区別するために閾値を設定することが困難であることがわかる。

次に注目するのは、他人に対する類似度である。最大固有値の場合には、顔入力に対する他人の類似度が最も大きいケースは 0.853 (m002 の辞書に対する m011real) である (写真入力に対する他人の類似度が最も大きいケースは 0.780 (m010 の辞書に対する m007photo) である)。一方、第 3 固有値を用いた場合には、顔で 0.617 (m004 の辞書に対する m003real) である (写真では 0.414 (m006 の辞書に対する m003photo) である)。したがって、最大固有値でも第 3 固有値でも、他人との区別をする閾値は容易に設定可能である。

5 考察

ここで仮定したのは、次の 3 点である。

- 観測方向の変化が微小であるため部分空間で近似できること。
- 観測方向の変化は横と縦の回転 (図 3 の Ω_x と Ω_y) であること。
- 類似性の評価に最大ノルムの考えから第 3 固有値を用いること。

ここで、理解を助けるために、回転が 1 方向 (たとえば横方向) のみの場合を考えてみよう。この場合は、

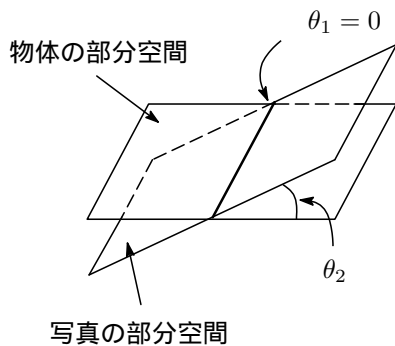


図 7: 回転が 1 方向の場合

部分空間は 2 次元となる。正面から見た画像は、両者とも同じであるので、最大固有値は 1、すなわち、最小正準角は 0 度となるはずである。一方、回転するにしたがって、両者の見え方は異なってくるので、両者の部分空間は完全に一致している訳ではない。この様子を図 7 に示す。

このとき、両方の部分空間の違いをどのように定義するのが良いかということを考えてみる。これは、あくまで、どのように評価したいのかという問題であり、どのような 3 次元認識を行ないたいと思うのかという問題である。その場面によって、妥当な方法は変化する。本研究で採用したのは、図の θ_2 を使うという案であった。

6 むすび

部分空間の間の正準角を利用することにより、3 次元パターン間の類似度を定義するという方式を提案し、2 種類の実験によって有効性を確認した。

本論文では、観測方向の変化が微小であるという仮定のもとに、3 次元物体を観測した画像の分布が 3 次元で近似できるとした。また、3 次元物体の観測方向の変化として、横と縦（図 3 の Ω_x と Ω_y ）を想定した。（実際に、顔を動かすように指示すると、ほとんどが、この回転運動になる。）類似性の評価は、文献 [8] をヒントに、最大ノルムの考え方によった。

実際の場面では、観測方向が大きく変化したり、 Ω_z の回転もあり得るし、類似性の評価も、文献 [10] で提案されたように、複数の正準角の使い方には、いくつかの他の候補もあり得る。これらの場合に、本論文で

提案した方式をどのように拡張していけばよいのかということが、将来の研究課題となる。さらに、制約空間 [6] に射影した場合にも、本論文で提案した方法が有効であるか否かも残されて問題である。

また、比較的少ないデータでの評価実験であるため、100 名程度以上のデータに対して、エラー率の変化も詳細に調べる必要があるものと思われる。

参考文献

- [1] 村瀬 洋, シュリー・ナイヤー, “2 次元照合による 3 次元物体認識”, 信学論 (D-II), vol. J77-D-II, no. 11, pp.2179-2187, 1994.
- [2] 前田 賢一, 渡辺 貞一, “局所的構造を導入したパターン・マッチング法”, 信学論 (D), vol. J68-D, no. 3, pp.345-352, 1984.
- [3] K. Maeda, K. Fukui and O. Yamaguchi, “Towards 3-Dimensional Pattern Recognition”, Proc. of SPR04, 2004
- [4] エルッキ・オヤ著, 小川 英光, 佐藤 誠 訳, “パターン認識と部分空間法”, 産業図書, 1986.
- [5] 山口 修, 福井 和広, 前田 賢一, “動画像を用いた顔認識システム”, 信学技報, PRMU97-70, pp. 17-24, 1997
- [6] 福井 和広, 山口 修, 鈴木 薫, 前田 賢一, “制約相互部分空間法を用いた環境変動にロバストな顔認識 - 照明変動を抑える制約部分空間の学習 -”, 信学論 (D-II), vol. J82-D-II, no. 4, pp. 613-620, 1999
- [7] F. Chatelin, “行列の固有値”, 伊理 正夫, 伊理 由美 訳, シュプリングラー・フェアラーク東京, 1993
- [8] E. Oja and J. Parkkinen, “On Subspace Clustering,” Proc. of ICPR '84, pp.692-695, 1984
- [9] 福井 和広, 山口 修, “形状抽出とパターン照合の組合せによる顔特徴点抽出”, 信学論 (D-II), vol. J80-D-II, No.8, pp2170-2177, 1997

- [10] K. Fukui and O. Yamaguchi , “Face Recognition Using Multi-viewpoint Patterns for Robot Vision” , Proc. of ISRR03, 2003

表 3: 類似度の分布

顔 (最大固有値)	m001	m002	m003	m004	m005	m006	m007	m008	m009	m010	m011
m001real	<i>0.995</i>	0.717	0.673	0.796	0.709	0.652	0.744	0.651	0.740	0.760	0.750
m002real	0.696	<i>0.994</i>	0.683	0.763	0.737	0.772	0.752	0.792	0.805	0.814	0.809
m003real	0.676	0.649	<i>0.993</i>	0.835	0.713	0.755	0.644	0.647	0.738	0.626	0.622
m004real	0.794	0.741	0.833	<i>0.992</i>	0.766	0.722	0.697	0.701	0.746	0.748	0.738
m005real	0.714	0.691	0.669	0.684	<i>0.993</i>	0.773	0.727	0.751	0.700	0.691	0.702
m006real	0.663	0.762	0.747	0.740	0.762	<i>0.998</i>	0.581	0.702	0.734	0.673	0.736
m007real	0.699	0.742	0.626	0.711	0.690	0.561	<i>0.995</i>	0.640	0.696	0.767	0.652
m008real	0.591	0.824	0.698	0.696	0.726	0.725	0.632	<i>0.990</i>	0.824	0.696	0.748
m009real	0.750	0.783	0.707	0.722	0.736	0.755	0.685	0.798	<i>0.995</i>	0.806	0.759
m010real	0.725	0.821	0.630	0.696	0.741	0.664	0.819	0.657	0.783	<i>0.993</i>	0.757
m011real	0.730	0.853	0.663	0.712	0.738	0.702	0.728	0.789	0.766	0.784	<i>0.970</i>
顔 (第3固有値)	m001	m002	m003	m004	m005	m006	m007	m008	m009	m010	m011
m001real	<i>0.947</i>	0.306	0.533	0.520	0.522	0.376	0.427	0.348	0.538	0.525	0.355
m002real	0.295	<i>0.947</i>	0.209	0.177	0.267	0.340	0.446	0.541	0.544	0.408	0.417
m003real	0.478	0.203	<i>0.951</i>	0.617	0.470	0.408	0.350	0.425	0.409	0.527	0.317
m004real	0.505	0.311	0.589	<i>0.915</i>	0.519	0.287	0.428	0.350	0.308	0.386	0.321
m005real	0.473	0.226	0.505	0.525	<i>0.927</i>	0.425	0.484	0.542	0.435	0.329	0.279
m006real	0.459	0.431	0.468	0.358	0.479	<i>0.959</i>	0.133	0.447	0.310	0.351	0.321
m007real	0.292	0.506	0.241	0.252	0.432	0.119	<i>0.940</i>	0.332	0.218	0.212	0.320
m008real	0.359	0.506	0.379	0.411	0.540	0.345	0.363	<i>0.929</i>	0.634	0.470	0.280
m009real	0.519	0.365	0.434	0.378	0.298	0.219	0.247	0.590	<i>0.936</i>	0.479	0.459
m010real	0.525	0.629	0.538	0.506	0.409	0.310	0.425	0.454	0.666	<i>0.959</i>	0.428
m011real	0.369	0.589	0.448	0.433	0.365	0.285	0.343	0.421	0.604	0.539	<i>0.905</i>
写真 (最大固有値)	m001	m002	m003	m004	m005	m006	m007	m008	m009	m010	m011
m001photo	<i>0.809</i>	0.547	0.577	0.603	0.425	0.489	0.464	0.564	0.637	0.534	0.451
m002photo	0.611	<i>0.829</i>	0.551	0.625	0.501	0.565	0.649	0.638	0.735	0.748	0.578
m003photo	0.605	0.612	<i>0.958</i>	0.747	0.569	0.707	0.533	0.687	0.635	0.496	0.461
m004photo	0.705	0.707	0.783	<i>0.946</i>	0.697	0.700	0.637	0.660	0.729	0.644	0.623
m005photo	0.497	0.450	0.450	0.575	<i>0.776</i>	0.580	0.496	0.554	0.551	0.442	0.422
m006photo	0.573	0.644	0.725	0.622	0.719	<i>0.965</i>	0.522	0.672	0.645	0.532	0.616
m007photo	0.663	0.692	0.527	0.622	0.609	0.443	<i>0.961</i>	0.581	0.559	0.780	0.608
m008photo	0.653	0.667	0.656	0.706	0.710	0.612	0.599	<i>0.888</i>	0.723	0.584	0.531
m009photo	0.698	0.748	0.636	0.656	0.651	0.719	0.643	0.740	<i>0.956</i>	0.698	0.655
m010photo	0.613	0.762	0.541	0.605	0.613	0.521	0.733	0.609	0.682	<i>0.960</i>	0.632
m011photo	0.685	0.627	0.587	0.662	0.649	0.531	0.594	0.632	0.681	0.551	<i>0.779</i>
写真 (第3固有値)	m001	m002	m003	m004	m005	m006	m007	m008	m009	m010	m011
m001photo	<i>0.218</i>	0.093	0.279	0.349	0.044	0.245	0.065	0.182	0.243	0.255	0.159
m002photo	0.205	<i>0.483</i>	0.167	0.143	0.107	0.126	0.186	0.178	0.280	0.241	0.220
m003photo	0.216	0.192	<i>0.398</i>	0.337	0.324	0.414	0.179	0.173	0.285	0.298	0.175
m004photo	0.152	0.112	0.380	<i>0.616</i>	0.160	0.246	0.107	0.202	0.255	0.286	0.228
m005photo	0.063	0.056	0.221	0.208	<i>0.055</i>	0.308	0.053	0.180	0.119	0.109	0.070
m006photo	0.116	0.118	0.203	0.273	0.306	<i>0.598</i>	0.140	0.196	0.211	0.251	0.147
m007photo	0.128	0.278	0.128	0.163	0.273	0.288	<i>0.607</i>	0.200	0.249	0.435	0.148
m008photo	0.140	0.148	0.236	0.198	0.141	0.266	0.191	<i>0.467</i>	0.324	0.263	0.123
m009photo	0.250	0.364	0.179	0.310	0.179	0.307	0.106	0.204	<i>0.587</i>	0.347	0.372
m010photo	0.182	0.134	0.156	0.251	0.181	0.385	0.225	0.218	0.273	<i>0.459</i>	0.216
m011photo	0.173	0.134	0.137	0.237	0.087	0.188	0.134	0.166	0.293	0.238	<i>0.295</i>