

言語メタデータに基づく画像コンテンツの検索

Image Contents Retrieval based on Linguistic Meta-Data

小泉 大地[†] 大西 正裕[†] 柘植 覚[‡] 獅々堀 正幹[‡] 北 研二[§]
Daichi Koizumi Masahiro Ohnishi Satoru Tsuge Masami Shishibori Kenji Kita

1. はじめに

画像データベース中の各画像に対し言語的なメタ情報を付与することにより、画像に対する意味的検索を行うことが可能となる。しかし、データベースが大規模な場合、人手によりすべてのデータにメタ情報を付与することは困難であり、自動的あるいは半自動的に付与する手法が必要となる。

本稿では、混合ガウス分布モデル (Gaussian Mixture Model; GMM) を用いて、画像中の各領域に適切な言語メタデータ (キーワード) を付与する手法を提案する。また、風景画像を用いた実験により、提案した手法の有効性を評価する。

2. 混合ガウス分布モデル

2.1 混合ガウス分布

混合ガウス分布 $P(\mathbf{x}_n|\theta)$ は任意の分布を複数のガウス分布 $P(\mathbf{x}_n|k, \theta)$ の線形結合で近似するものであり、以下のように定義される。

$$P(\mathbf{x}_n|\theta) = \sum_{k=1}^K P(\mathbf{x}_n|k, \theta)\pi_k \quad (1)$$

$$P(\mathbf{x}_n|k, \theta) = \frac{1}{\sqrt{(2\pi)^M |R_k|}} \exp\left\{-\frac{1}{2}(\mathbf{x}_n - \mu_k)^T R_k^{-1}(\mathbf{x}_n - \mu_k)\right\} \quad (2)$$

なお、 \mathbf{x}_n は M 次元の特徴ベクトルであり、モデルのパラメータ $\theta = \{\pi_k, \mu_k, R_k, K\}$ は以下の通りである。

π_k : ガウス分布 $P(\mathbf{x}_n|k, \theta)$ の重み ($\sum_k \pi_k = 1$)

μ_k : M 次元平均ベクトル

R_k : $M \times M$ 共分散行列

k : ガウス分布 $P(\mathbf{x}_n|k, \theta)$ の混合数

多次元特徴ベクトルの系列 $\mathbf{x} = \{\mathbf{x}_n\}_{n=1}^N$ に対し、対数尤度は以下のように定義される。

$$\log P(\mathbf{x}|K, \theta) = \sum_{n=1}^N \log \left(\sum_{k=1}^K P(\mathbf{x}_n|k, \theta)\pi_k \right) \quad (3)$$

[†]徳島大学 大学院工学研究科

[‡]徳島大学 工学部

[§]徳島大学 高度情報化基盤センター & 工学部

2.2 パラメータ推定

混合ガウス分布 $P(\mathbf{x}_n|\theta)$ のパラメータ $\theta = \{\pi_k, \mu_k, R_k, K\}$ は、EM アルゴリズムに基づく繰り返し計算により推定することができる。EM アルゴリズムでは、以下の $Q(\theta; \bar{\theta})$ 関数が最大となる更新値 $\bar{\theta} = \{\bar{\pi}_k, \bar{\mu}_k, \bar{R}_k, \bar{K}\}$ を求める [1]。

$$\begin{aligned} Q(\theta; \bar{\theta}) &= \sum_{k=1}^K \bar{N}_k \left\{ -\frac{1}{2} \text{Tr}[\bar{R}_k \bar{R}_k^{-1}] \right. \\ &\quad - \frac{1}{2} (\bar{\mu}_k - \mu_k)^T \bar{R}_k^{-1} (\bar{\mu}_k - \mu_k) \\ &\quad - \frac{M}{2} \log(2\pi) - \frac{1}{2} \log(|R_k|) + \log(\pi_k) \left. \right\} \\ &\quad - \frac{1}{2} L \log(NM) \end{aligned}$$

$$\bar{N}_k = \sum_{n=1}^N P(k|\mathbf{x}_n, \bar{\theta}) \quad (\sum_k \bar{N}_k = N)$$

$$\bar{\pi}_k = \frac{\bar{N}_k}{N}$$

$$\bar{\mu}_k = \frac{1}{\bar{K}_k} \sum_{n=1}^N \mathbf{x}_n P(k|\mathbf{x}_n, \bar{\theta})$$

$$\bar{R}_k = \frac{1}{\bar{K}_k} \sum_{n=1}^N (\mathbf{x}_n - \bar{\mu}_k)(\mathbf{x}_n - \bar{\mu}_k)^T P(k|\mathbf{x}_n, \bar{\theta})$$

3. GMMを用いた言語メタデータ生成手法

本節では、GMMを用いた画像の領域に対するキーワード付与手法を提案する。以下では、GMMの学習処理およびGMMによる言語メタデータ付与処理について述べる。

3.1 GMMの学習処理

原画像を複数の領域に分割し、各領域に対し適当と思われるキーワードを手で付与する。その後、キーワードごとに領域を分類し、領域内のすべての画素を対象に xy 座標値および LUV 色情報から成る 5 次元の特徴ベクトルを作成する。この際、各次元の特徴ベクトルの分散を考慮し、 xy 座標値および LUV 色情報にはあらかじめ正規化を施しておく。正規化した特徴ベクトルをキーワードごとに 1 つの系列にまとめ、各キーワードに対する GMM を作成する。なお、GMM の推定には参考文献 [1] の手法を用いたが、この手法

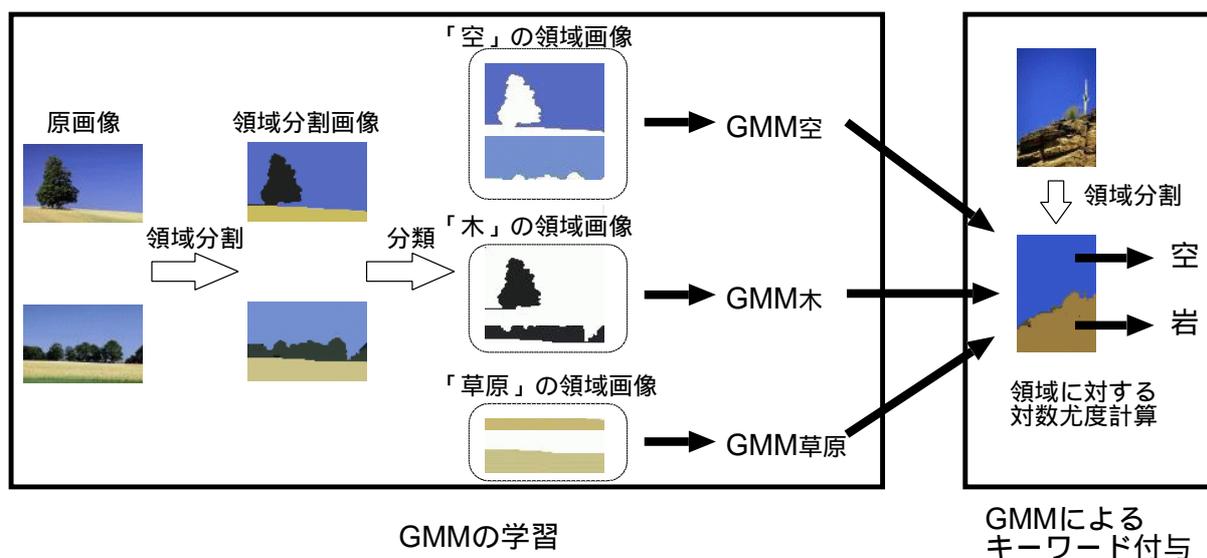


図 1: GMM を用いた領域画像への言語メタデータ付与処理

では、MDL 原理に基づき GMM の混合数を自動的に推定している。

図 1 の例では、原画像がそれぞれ 3 つの領域に分割され、各領域に対し「空」「木」「草原」といったキーワードが付与されている。この場合「空」「木」「草原」の各領域から得られる特徴ベクトルの系列から、それぞれのキーワードに対する GMM を作成する。

3.2 GMM による言語メタデータ付与処理

GMM 学習処理と同様に、入力画像を領域分割後、各領域の特徴ベクトルを作成する。式 (3) から、各キーワードの GMM に対する領域の特徴ベクトル系列の対数尤度を求め、値が大きいものをその領域のキーワード候補とする。

図 1 の例では、まず入力画像から分割された領域を対象に特徴ベクトル系列を作成する。次に作成した特徴ベクトル系列と学習したキーワードの GMM に対する対数尤度を求め、最も大きい値となる「空」をキーワード候補として出力する。言語メタデータ付与実行結果例を図 2 に示す。

4. 評価実験

本節では、提案手法の有向性を調べるために学習用データと評価用データを用意し比較実験を行った。以下では、実験概要および実験結果について述べる。

4.1 実験概要

評価実験では、Corel 画像データベース [2] から抽出した風景のカラー写真画像 500 件を用い、このうち、400 件を学習用データ、100 件を評価用データとした。

まず、各画像を平均移動アルゴリズム (mean shift algorithm) [3] を用いて、自動的に複数の領域に分割した。なお、領域分割の際には色情報のみを用いており、1 画像当りの領域数は平均 12 個と比較的粗い分割となっている。その後、各領域に対して適当と思われるキーワード (例：海、山、草原等、総数 67 個) を人手により付与した [4]。多くの場合は 1 つの領域に対し 1 つのキーワードのみが付与されているが、領域分割の仕方によっては、1 つの領域に複数のキーワードが付与されているものも少数ながら存在する。

次に、学習用画像 400 件から得られた領域 (領域総数 4836 個) から、画素の xy 座標値および LUV 色情報から成る 5 次元の特徴ベクトルを作成した。作成した特徴ベクトルを領域に付与されたキーワードごとに分類し、各キーワードに対する GMM を作成した。また、混合数の最大値を 8, 16, 32 と変化させ実験を行った。

キーワードごとに推定された GMM を用いて、式 (3) により、評価用画像中の領域に対する対数尤度を求める。本実験では、対数尤度が大きいもの上位 5 件から領域に対する候補キーワードを生成した。

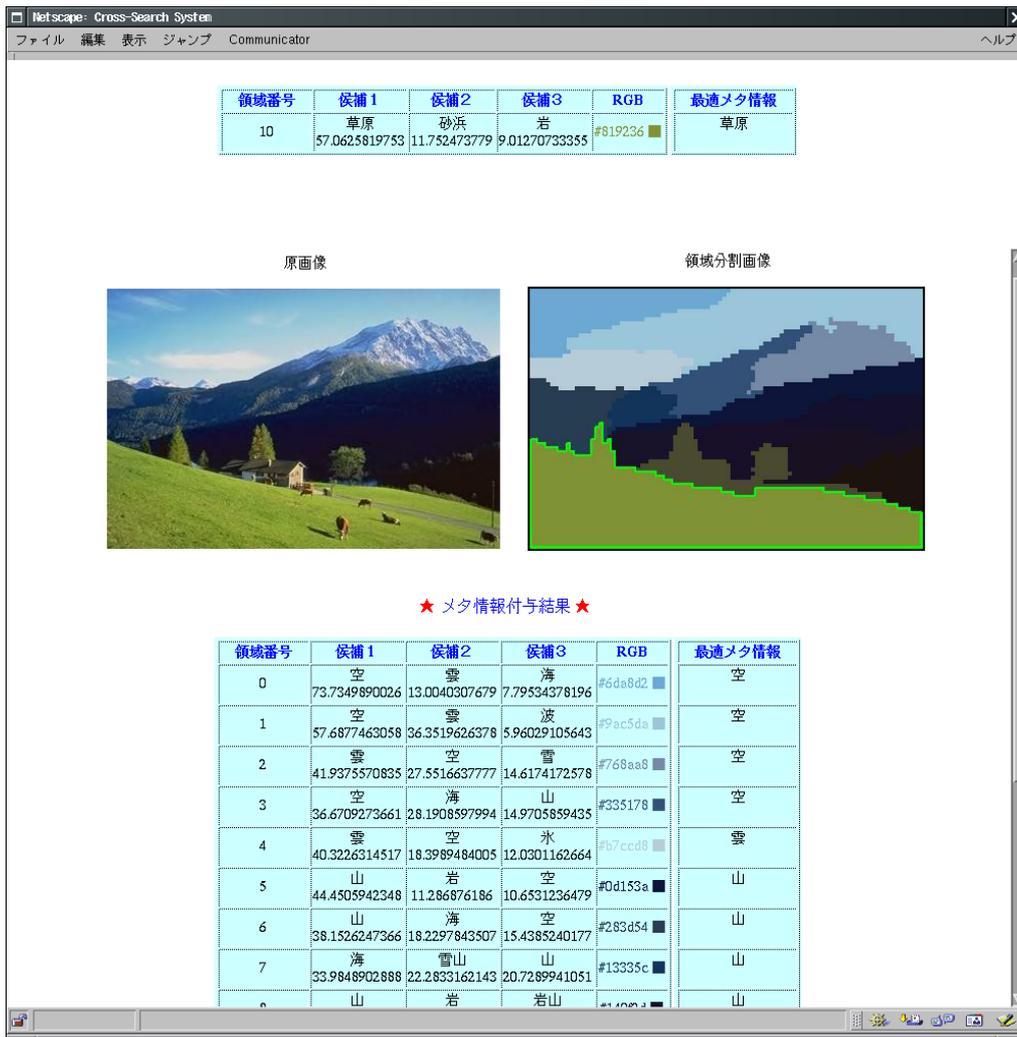


図 2: 言語メタデータ付与実行結果例

表 1: 言語メタデータの累積正解率

順位	混合数 8	混合数 16	混合数 32
1	28.09%	30.38%	31.97%
2	42.30%	44.51%	47.43%
3	53.67%	55.88%	58.33%
4	62.19%	64.87%	67.56%
5	68.74%	71.42%	73.09%

4.2 実験結果および考察

評価用画像 100 件から得られる領域 (領域総数 1267 個) に対し, 人手で付与した正解キーワードと, 提案手法により付与した上位 x 件 ($1 \leq x \leq 5$) 内に含まれ

る候補キーワードとを比較することで累積正解率計算した. 結果を表 1 に示す.

実験結果より, 候補キーワード上位 5 件までの累積正解率は 70% を達成した. また, 混合数の増加により累積正解率も弱冠上昇した.

誤りの例として, 「太陽」, 「道路」, 「石」など学習データ数が少ないキーワードが確認できた. これらキーワードの学習データ数は 3000 以内となっており, 正解例キーワードの学習データ数と比べ $1/5 \sim 1/10$ の学習データ数となっている. 以上より, 提案手法での各キーワードの学習データ数はある程度用意する必要があると考えられる.

次に, 混合数 32 における代表的なキーワードに対する混同行列を調査した. 結果を表 2 に示す. 縦軸

表 2: 代表的なキーワードに対する混同行列

	空	雲	水辺	海岸	木	岩	岩山	雪山	紅葉	波	森	雪	野原	雪景色	川
空	178	29	1	0	5	3	4	6	0	0	0	4	1	4	4
雲	8	41	3	3	0	1	12	3	2	0	0	2	0	1	0
水辺	0	0	4	4	0	0	0	2	0	0	0	3	1	1	1
海岸	0	0	7	8	1	1	3	3	0	1	2	1	4	0	1
木	1	1	0	0	8	2	0	0	0	0	15	0	3	0	2
岩	0	1	1	9	7	32	28	2	10	0	19	0	15	0	6
岩山	0	3	0	0	1	6	24	1	2	0	0	0	2	0	0
雪山	6	3	1	1	0	0	0	2	0	0	0	4	0	0	0
紅葉	0	0	0	0	0	1	4	0	4	0	0	0	3	0	0
波	2	4	2	0	0	0	0	3	0	2	0	3	0	2	0
森	0	0	0	1	8	2	0	0	2	0	4	0	1	0	1
雪	4	2	1	1	0	1	1	2	0	1	0	3	0	4	0
野原	0	0	0	1	3	5	3	0	2	0	11	0	10	0	3
雪景色	1	1	0	1	0	1	1	0	0	0	0	0	0	2	0
川	0	1	1	0	0	0	0	1	0	1	1	2	0	1	4

は正解キーワード、横軸は候補キーワードであり、正解キーワードに対する候補キーワードの数をカウントする。正答数が高かったキーワードに対し、正解キーワード以外にどのような候補キーワードが付与されているかを調べたところ、「雲」に対する領域に「空」というキーワードを付与したものの、「岩」に対する領域に「岩山」というキーワードを付与したものが確認できた。一般に、領域に対しキーワードを付与する場合、付与する人間によりキーワードは変化する。そのため、上記の例のように、関連語となるキーワードはあらかじめ何らかの形にまとめておく必要があると考えられる。

また、表 2 が対角状になっていることより、候補キーワードの正答数が多いことが確認できた。特に「空」、「海」は 178, 44 と高い正答数であった。一般に「空」、「海」は色情報の変化が少ない。また、「空」は画像上部に存在するなどのキーワードの特性から座標値の変化も少なくなる。そのため、高い正答数であったと考えられる。

5. まとめ

本稿では、キーワードごとの GMM を作成することにより、画像中の各領域に適したキーワードを付与する手法を提案した。また風景画像を用いた評価実験

により、提案手法を用い付与したキーワード上位 5 件までの累積正解率は 70% を達成した。

今後の課題として、学習データ数の変化による精度比較、また各キーワード間の相関関係を考慮した言語メタデータ生成手法の提案が挙げられる。

参考文献

- [1] C. A. Bouman. Cluster: An unsupervised algorithm for modeling gaussian mixtures. 1997. <http://www.ece.purdue.edu/~bouman>.
- [2] Corel Gallery homepage. <http://www.corel.co.jp/>.
- [3] D. Comaniciu and P. Meer. Robust analysis of feature spaces: Color image segmentation. *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 750–755, 1997.
- [4] 小泉大地, 柘植覚, 獅々堀正幹, 北研二. テキストと画像のクロスメディア情報検索に向けた画像キーワード登録システムの開発. 情報処理学会データベースシステム研究会資料, No. DBS-127, pp. 105–112, 2002.