文書中の地物画像の内容を説明する言語的記述の Web からの抽出

Extracting Visual Description on the Web as the Linguistic Alternative to Image of Geographic Feature in Digital Document

服部 峻♥　手塚 太郎♦
田中 克己♠

Shun HATTORI　Taro TEZUKA
Katsumi TANAKA

近年，誰もが，どんな計算機環境からでも，Web 文書などのデジタル情報に対してユニバーサル・アクセス可能にする技術の一つとして，画像メディアを出力することができない非視覚的ブラウザのために，画像を含むマルチメディアな文書から，画像を含まないテキスト版に変換する技術が必要とされている．本論文では，この小問題として，デジタル文書中の画像，特に，著名な地物（建築物）を含む画像に対して，その画像に写っている内容を説明する言語メディア表現を自動生成するために，地物の外観情報を，その文書中だけでなく Web 全体から抽出して来る手法について提案する．文書中の各画像に対して，その周辺テキストや文書タイトルなどの文脈に基づいて，画像内の地物の名称を推定し，その上で，その地物の外観に関する言語的な記述を Web マイニングにより抽出することで，対象の画像の内容を表す代替テキストとして対応付けることができる．

Recently, “ubiquitous society” requires techniques to universally access digital information. One of the more important techniques allows us to access the Web via non-graphical browsers such as text or audio browsers. In this paper, we propose a novel method to extract visual description as the linguistic alternative to a photograph of geographic feature in a digital document by mining the Web, in an effort to convert the original multimedia version of the digital document into the text-only version by replacing each photograph with its visual description. Our target system first identifies the names of geographic features which are taken in each photograph, based on its context such as its surrounding text or the title of the digital document with it, and then extracts visual description for its each identified name of geographic feature by mining the Web as well as the content of the digital document with it.

♥ 学生会員　京都大学大学院 情報学研究科 博士後期課程
hattori@dl.kuis.kyoto-u.ac.jp
♦ 正会員　京都大学大学院 情報学研究科 社会情報学専攻
tezuka@dl.kuis.kyoto-u.ac.jp
♠ 正会員　京都大学大学院 情報学研究科 社会情報学専攻
tanaka@dl.kuis.kyoto-u.ac.jp

1. はじめに

近年，ユビキタス社会を実現するための大きな課題として，誰もが，どんな計算機環境からでも，デジタル情報に対してユニバーサル・アクセス可能にする技術が求められている．その一つとして，画像メディアを出力できないテキストブラウザ[1]や音声ブラウザ[2]といった非視覚的ブラウザからでもWeb文書にアクセスすることを可能にするために，画像メディアが混合した元々のマルチメディアなWeb文書を，言語的記述のみのテキスト版に変換する技術が必要である．

従来の非視覚的ブラウザの多くは，Web画像の代わりに，そのIMG要素のALT属性の代替テキスト，SRC属性の画像ファイル名，その画像がリンクアンカーであればリンク先のWeb文書のタイトルやURLなどのテキスト情報で置換することで，マルチメディアなWeb文書をテキスト版に変換している．しかしながら，変換対象のWeb文書において画像自体が著者の伝えたい主要なコンテンツである場合には，従来の単純な置換方法では，元々のマルチメディアなWeb文書から視覚的に得られる情報とは程遠い価値のテキスト版しか得ることができない．一般の視覚的ブラウザからと非視覚的ブラウザからとでは，同一のWeb文書にアクセスしたとしても得られる情報に大きな差があり，依然としてデジタルデバイドの問題を払拭できていない．

この問題に対して，Web画像の内容を説明する言語的な記述として，その画像中のオブジェクトの名称やオブジェクト間の位置関係といった高次の意味的な情報や，そのオブジェクトの色名，大きさ，形状名などの高次の視覚的な情報を対応付けることができれば，画像メディアと言語メディアは表現能力などの特長が元々異なるため完全に等価な表現メディア変換は不可能であるとしても，Web文書の一層のマルチメディア化に伴うデジタルデバイドの問題を緩和することができると我々は考える．

画像の内容を表す言語的な記述を求める研究は，画像認識や画像理解の分野で古くから盛んに行われているが，光学的文字認識（OCR）や特定の対象領域に限定したオブジェクト認識などを除き，一般のオブジェクトを認識する実用的な手法は未だ考案されていない[3][4]．一方で，画像検索の分野では，Web画像に対して，関連キーワードや説明文などのテキスト情報を索引付けする様々な手法が提案されているが，非視覚的ブラウザからのユニバーサル・アクセスを実現するための画像の代替テキストとして，十分に適当なテキスト情報を抽出できている研究例は見当たらない[5][6]．

そこで，本論文では，この小問題として，デジタル文書中の画像，特に，著名な地物（建築物）を含む画像に対して，その画像に写っている内容を説明する言語メディア表現を自動生成するために，地物の外観情報をWeb全体から抽出する手法について提案する．文書中の各画像に対して，その周辺テキストや文書タイトルなどの文脈に基づいて，画像内の地物の名称を推定し，その上で，その地物の外観に関する言語的な記述をWebマイニングにより抽出することで，対象の画像の代替テキストとして対応付けることができる．

本論文の以下の構成を示す．2章では，対象の地物の全体としての外観情報をWebから抽出する手法について提案し，抽出例も示す．3章では，対象の地物を構成する要素の名称，及び，その構成要素毎の外観情報をWebから抽出する手法について提案し，各々の抽出例も示す．最後に，5章で本論文をまとめ，今後の研究課題についても述べる．

2. 地物の全体的な外観情報の抽出

本章では，対象の地物が全体として，典型的にどのように見られているかを表す外観情報を，Web マイニングにより抽出する手法について述べる．多くの地物（建築物）に共通する外観情報としては，色，大きさ，高さ（階数），形状，材質などが考えられる．全く同一の地物を見たとしても，個々人に依って印象は異なり，それを表現する仕方も千差万別である．逆に，ある地物の外観情報として呈示された記述から連想するイメージも個々人に依って異なるため，地物画像の代替テキストとして，その地物の外観情報を呈示したとしても，完全に等価な表現メディア変換とはならない．しかしながら，そもそも同一の地物画像から得られる情報が個々人に依って異なることと比べれば，言語的な説明の方が受け取り方の変動が小さいため，地物の外観情報は，その地物の画像の代替テキストとして有用性があると考える．

2.1 抽出手法

対象の地物の名称に対して，抽出したい外観属性毎に，次の二つのステップを実行することによって，各候補値の相応度が評価され，より相応しい外観属性の値が求められる．

Step 1: 外観属性の候補値の取得
対象の地物名と抽出したい外観属性の表現テンプレートを基に構成した検索質問を文書検索エンジンに投げ，その検索結果のスニペット（要約部）を解析することで，抽出したい外観属性の値の候補語を得る．外観属性として色名に対しては「色の」，材質・様式に対しては「造りの」，階数に対しては「階建ての」，屋根の葺き方に対しては「葺きの」といった表現テンプレートを用いる．

Step 2: 外観属性の候補値の相応度評価
外観属性毎の各候補値 v に対して，文書検索エンジンの検索件数を用いて，対象の地物名 g への相応度 $S_g(v)$を次式により評価し，他よりも十分に大きな相応度となった候補値を外観属性毎の推定値として採択する．

$$S_g(v) = \frac{df(g \wedge v)}{df(v)}$$

但し，df(q)は，文書検索エンジンのコーパス中で，検索質問 q に合致する文書の総数を表すものとする．

図 1: 京都の祇園にある八坂神社の西楼門の写真の一例
Figure 1: Example of Photograph of Nishi-Romon of Yasaka-Jinja in Gion, Kyoto

2.2 抽出例

地物の外観情報の抽出例として，本論文では，以下のような外観的な特徴を持つ京都祇園の「八坂神社」を対象とする．

・西に二層の朱色の西楼門（図1），南には大きな石鳥居や朱色の南楼門，中央の境内には祇園造りの本殿が在る．
・敷地全体が緑色の森に包まれている．

本節では，「八坂神社」の全体としての外観情報を抽出する例を示す．色名の適合解としては「朱色」が想定され，類似する「赤色」や，森の「緑色」なども相応度が高く評価されることが期待される．材質・様式としては「石造り」や「祇園造り」が，階数としては不定や「2階建て」が想定される．

「八坂神社」の色情報を抽出するためには，まず，色名の候補語を得る．["八坂神社" AND "色の"]という検索質問を Google のブログ検索エンジン[7]で検索した結果のスニペットから「色の」の直前の名詞を切り出し，色名辞書に登録されている語だけを候補語とする．次に，色名の各候補に対して，「八坂神社」という地物の色情報としての相応度を算出すると表 1 のようになる．但し，df(q)には，Google のブログ検索エンジンの検索件数を用いている．「朱色」は他の色名よりも約 10 倍以上大きな相応度となっており期待通りの結果である．一方で，「朱色」に類似する「赤色」や「オレンジ色」，或いは，森の「緑色」の値は，他よりは幾分大きいながら有意な差があるとまでは言えず，何らかの閾値によって「八坂神社」の色情報として採択するのは困難である．もちろん，「赤色」や「オレンジ色」については，「朱色」が抽出できさえすれば，色名の間の類似関係を予め辞書化しておいたり，或いは，色間の類似度[8]を計算したりすれば，「朱色」の類似色として呈示することは可能である．しかしながら，「八坂神社」を包んでいる森の「緑色」については，現時点では採択する術が無く，今後，改善する余地が残る．

表 1: 地物名（g = "八坂神社"）の外観語（v）の相応度
Table 1: Suitability of Visual Description (v) for Geographic Feature's Name (g = "Yasaka-Jinja")

v	df(g∧v)	df(v)	$S_g(v)$
白色	8	73232	$0.10924 \cdot 10^{-3}$
黒色	3	44653	$0.06718 \cdot 10^{-3}$
赤色	12	70742	$0.16963 \cdot 10^{-3}$
朱色*	**41**	**18440**	**$2.22343 \cdot 10^{-3}$**
ピンク色	14	97531	$0.14354 \cdot 10^{-3}$
オレンジ色	19	84265	$0.22548 \cdot 10^{-3}$
黄色	46	254070	$0.18105 \cdot 10^{-3}$
緑色	22	120612	$0.18240 \cdot 10^{-3}$
青色	10	72010	$0.13887 \cdot 10^{-3}$
藍色	2	18406	$0.10866 \cdot 10^{-3}$
紫色	10	67797	$0.14750 \cdot 10^{-3}$
祇園造り*	**3**	**3**	**1.00000**
寝殿造り	4	192	0.02083
数奇屋造り	4	324	0.01235
出桁造り	1	115	0.00870
レンガ造り	11	3801	0.00289
2 階建て	17	19261	$0.88261 \cdot 10^{-3}$
3 階建て	5	10174	$0.49144 \cdot 10^{-3}$
4 階建て	2	3979	$0.50264 \cdot 10^{-3}$
10 階建て	2	1505	$1.32890 \cdot 10^{-3}$

3. 地物の構成要素毎の外観情報の抽出

文書中の地物画像の周辺テキストなどの文脈には，その画像に写っている地物の固有名称に関しては記述されている場合が多いが，その地物の特にどの構成要素について撮られた写真であるかまでは記述されていない場合もある．例えば，前章の図1の写真に対して，その文脈に「八坂神社」という地物の固有名称が現れることは多いが，「西楼門」という構成要素名が現れることは多くない．画像と言語的説明のこのようなギャップは，その画像を含む文書の著者にとって，地物名を読書に伝えることは重要であっても，その地物の特定の構成要素名を伝えることが特に重要ではない場合に起こり得る．しかし，一方で，その地物の特定の構成要素の画像を掲載しているということは，地物の全体としての外観情報よりはむしろ，地物の特定の構成要素の外観情報を伝えたいと考えてのはずである．多数の構成要素を持ち，その構成要素毎に外観情報が異なる地物も多く，表現メディア変換対象の地物画像に対して，その地物名しか特定せず，地物の全体としての外観情報を抽出する手法に比べ，その地物の構成要素名までを特定し，地物の特定の構成要素の外観情報を抽出する手法の方が，地物画像を代替する言語的記述として，より正確かつ詳細な情報となる．本章では，対象の地物に対して，外観として主に構成する要素の名称，及び，その構成要素毎の外観情報を，Webから抽出する手法について述べる．

3.1 地物の構成要素名の抽出

名称が与えられた地物の外観を主に構成する要素の名称のリストを Web マイニングにより抽出する手法としては，まず，[（地物名）の]という検索質問を文書検索エンジンに入力し，その検索結果のスニペットから「（地物名）の」に続く名詞句を形態素解析により切り出す手法が考えられる．しかしながら，日本語の文法における「の」という助詞は様々な用途で使用されるため，Google の文書検索エンジンなどで検索した結果を解析したとしても，地物の構成要素名とは関係のないキーワード（例えば「歴史」「祭」など）も多数含まれてしまい抽出精度は十分ではない．そこで，我々は，「（地物名）の」という記述が画像の周辺に現れる場合に，それに続く名詞句が対象の地物の構成要素名を表していることが多いことに着目し，対象の地物の名称に対して，次の二つのステップを実行することによって，地物の外観を構成する要素の名称をより精度良く抽出する手法を提案する．

Step 1: 地物の構成要素名の候補語の取得
対象の地物の名称に対して，[（地物名）の]という検索質問を文書検索エンジンではなく画像検索エンジンに入力し，その検索結果中のスニペットから「（地物名）の」に続く名詞句を形態素（係り受け）解析[9]により切り出し，対象の地物の外観を構成する要素の名称の候補語とする．

Step 2: 地物の構成要素名の候補語の相応度評価
地物の構成要素名の各候補語に対して，地物名と候補語を「の」で接続した[（地物名）の（構成要素名の候補語）]という検索質問を画像検索エンジンで検索した件数によって，対象の地物の構成要素名としての相応度を評価し，[（地物名）の]という検索質問の検索件数に比べて十分に大きな占有割合となる候補語を，対象の地物の外観を主に構成する要素の名称として採択する．

3.2 地物の構成要素毎の外観情報の抽出

地物名とそれを構成する要素の名称が与えられた場合に，その地物の全体としての外観情報ではなく，特定の構成要素の外観情報をWebから抽出する手法について述べる．

Step 1: 構成要素毎の外観修飾句の候補の取得
対象の地物名と構成要素名に対して，文書コーパスから「（地物名）の（修飾句）（構成要素名）」という形式の記述を収集し，「（地物名）の」と「（構成要素名）」の間の文字列を地物の構成要素の外観修飾句の候補とする．

Step 2: 構成要素毎の外観修飾句の候補の相応度評価
地物の構成要素毎の外観修飾句の各候補に対して，地物名と候補語を「の」で接続し直後に構成要素名を接続した[（地物名）の（外観修飾句の候補）（構成要素名）]という検索質問を文書検索エンジンで検索した件数によって，対象の地物の構成要素の外観修飾句としての相応度を評価し，全ての候補語の検索件数の総和に比べて十分に大きな占有割合となる候補語を，対象の地物の構成要素毎の外観修飾句（外観情報）として採択する．

3.3 抽出例

前章で用いた「八坂神社」に対して，まず，構成要素名の抽出例を示す．適合解としては，「楼門」「鳥居」「本殿」「森」などが想定される．Google の各種検索エンジンに対して[八坂神社の]という検索質問を入力した検索結果から候補語を切り出し，各々の相応度を算出すると，下の表 2 のようになる．上位 k 件の適合率 P_k は画像検索エンジンを用いた場合に確かに最も良いが十分とは言えない．再現率を上げるために少し下位まで含めただけでも，適合率が悪化する．

表 2: 地物「八坂神社」の構成要素名の相応度の比較
Table 2: Comparison of Suitabilities of Component for Geographic Feature's Name "Yasaka-Jinja"

種類	構成要素名の候補語と相応度（降順）
Web P_k= 1/5 1/10 3/15 6/20	**境内*** (792)，祭礼 (777)，近く (680)，祇園祭 (588)，前 (483)，祭神 (383)，夏 (298)，夏祭り (255)，祭 (186)，奥 (183)，**鳥居*** (174)，中 (174)，**本殿*** (166)，例祭 (126)，**能舞台*** (104)，例大祭 (100)，**西楼門*** (99)，神事 (96)，**狛犬*** (82)，こと (80)，歴史 (67)，節分祭 (63)，**正門*** (55)，創建 (46)，大晦日 (35)，**南楼門*** (30)，**森*** (30)，**枝垂れ桜*** (30)，**紅葉*** (26)，
Blog P_k= 1/5 2/10 5/15 8/20	近く (79)，前 (60)，**境内*** (55)，祭礼 (41)，奥 (27)，祇園祭 (22)，祭神 (17)，中 (15)，祭 (14)，**本殿*** (13)，神事 (13)，**鳥居*** (12)，夏祭り (12)，**能舞台*** (9)，**西楼門*** (8)，**正門*** (8)，**紅葉*** (8)，**狛犬*** (7)，例大祭 (6)，創建 (4)，こと (4)，節分祭 (3)，**枝垂れ桜*** (2)，例祭 (2)，**森*** (1)，歴史 (1)，大晦日 (1)，**南楼門*** (0)，夏 (0)，
画像 P_k= 3/5 4/10 6/15 8/20	祭礼 (71)，**境内*** (64)，**狛犬*** (42)，祇園祭 (39)，**鳥居*** (30)，節分祭 (30)，前 (26)，近く (25)，祭 (17)，**本殿*** (12)，**枝垂れ桜*** (11)，祭神 (11)，夏祭り (11)，中 (11)，**正門*** (9)，歴史 (9)，奥 (8)，**西楼門*** (7)，**紅葉*** (7)，大晦日 (7)，**森*** (6)，こと (6)，**能舞台*** (5)，例祭 (5)，例大祭 (5)，神事 (4)，**南楼門*** (3)，夏 (1)，創建 (0)，

表 3: 地物「八坂神社」の構成要素毎の外観修飾句
Table 3: Visual Description for Component of Geographic Feature's Name "Yasaka-Jinja"

要素名	構成要素毎の外観修飾句と相応度（降順）
境内 (23)	**小高い*** (2), 祇園祭 (2), **さして広くない*** (1), **傾斜のある*** (1), 拝殿前 (1), 隣清水寺 (1),
狛犬 (16)	**ブロンズの*** (1), **角のある*** (1), **にっこり*** (1), 石段脇の (1), 入り口にいる (1), 表の (1), 重文の (1), 神殿 (1), 人面 (1),
鳥居 (50)	**赤い*** (13), **石*** (5), 南の (5), 南門 (4), 南側の (3), **大きな*** (2), 南側 (2), 南 (2), **朱塗りの*** (1), **赤*** (1), 入口 (1), 正面 (1),
西楼門 (26)	正門である (5), **朱塗りの*** (4), **大きな*** (2), 正面玄関とも言える (1), 主門である (1), 顔となった (1), 西側にある (1),
南楼門 (18)	正門は (3), 西楼門から (2), 正門 (2), 本殿から (1), 鳥居をくぐり (1),
本殿 (21)	御 (4), 南楼門から (1), 西楼門から (1), 拝殿と (1), ご (1),
森 (3)	鎮守の (1), 奥の (1), ある (1)
能舞台 (5)	境内にある (2), 境内には古くから (1), 本殿の前にある (1), 山門と (1)

次に，「八坂神社」の適合な構成要素名に対して，Google の Web 検索エンジンに["八坂神社の*（構成要素名）"]というワイルドカード検索質問を入力した結果のスニペットを解析し，「八坂神社の」と「（構成要素名）」との間から構成要素毎の外観修飾句を抽出すると，上の表 3 のようになる．地物の構成要素毎の外観情報として適合な候補語は太字で表記している．但し，「境内」に対する「小高い」「傾斜のある」に関しては，八坂神社の境内の画像を晴眼者が閲覧したとしても視覚的に得ることが困難な情報であるため，その画像の内容を説明する代替テキストとしては適合解とは言えない．このような過多な外観情報を含めたとしても，適合率，再現率ともに未だ十分ではない．係り受け解析を用いて「（構成要素名）」を修飾している語句だけに限定したり，位置関係や方角などを表す語を含む候補を除去したりすれば，適合率の向上を見込める．しかし，「南楼門」「本殿」などに対しては，適合解が候補にさえ含まれていないため，外観修飾句の候補語を収集する手法を見直す必要がある．

4. まとめと今後の課題

本論文では，文書中の画像，特に，著名な地物（建築物）を含む画像に対して，その内容を説明する言語メディア表現を生成するために，地物の外観情報を Web 全体から抽出する手法について提案した．文書中の各画像に対して，その周辺テキストや文書タイトルなどの文脈に基づいて，画像内の地物の名称を推定し，その上で，その地物の外観に関する言語的な記述を Web マイニングにより抽出することで，対象の画像の代替テキストとして対応付ける．「八坂神社」という地物に対して実験を行った結果，地物の外観情報の抽出としては使用に堪える程度に精度良く抽出できたが，地物画像の代替テキストとしては未だ十分とは言えない．

今後の課題として，より大規模な評価実験を行い，地物の外観語としての相応度を量る尺度の改良や，地物の外観情報をより効率的に網羅する手法などについて検討を続ける．

[謝辞]

本研究の一部は，文部科学省 21 世紀 COE 拠点形成プログラム「知識社会基盤構築のための情報学拠点形成」（リーダー：田中克己，平成 14～18 年度），及び，文部科学省研究委託事業「知的資産の電子的な保存・活用を支援するソフトウェア技術基盤の構築」，異メディア・アーカイブの横断的検索・統合ソフトウェア開発（研究代表者：田中克己），及び，文部科学省科学研究費補助金特定領域研究「情報爆発時代に向けた新しい IT 基盤技術の研究」，計画研究「情報爆発時代に対応するコンテンツ融合と操作環境融合に関する研究」(研究代表者:田中克己, A01-00-02, 課題番号:18049041) によるものです．ここに記して謝意を表します．

[文献]

[1] Lynx, http://lynx.browser.org/ (2007).
[2] IBM Home Page Reader, http://www-06.ibm.com/jp/accessibility/soft/hpr.html (2007).
[3] 彌富仁, 萩原将文: "ファジー推論ニューラルネットワークを用いた風景画像からの知識抽出と認識," 電子情報通信学会 論文誌, Vol.82-D2, No.4, pp.685-693 (1999).
[4] 柳井啓司: "一般画像自動分類の実現へ向けた World Wide Web からの画像知識の獲得," 人工知能学会誌, Vol.19, No.5, pp.429-439 (2004).
[5] 相良直樹, 砂山渡, 谷内田正彦, "HTML テキストの重要文を用いた画像ラベリング手法," 電子情報通信学会 論文誌, Vol.J87-D1, No.2, pp.145-153 (2004).
[6] 竹内謹治, 黄瀬浩一, "類似画像とキーワードを利用した Web 画像の説明文抽出," 情報処理学会 研究報告「自然言語処理」, Vol.2006, No.1, pp.7-12 (2006).
[7] Google ブログ検索, http://blogsearch.google.co.jp/ (2007).
[8] John R. Smith and Shih-Fu Chang: "VisualSEEk: A Fully Automated Content-Based Image Query System," In Proc. of the 4th ACM International Conference on Multimedia, pp.87-98 (1996).
[9] Taku Kudo and Yuji Matsumoto: "Fast Methods for Kernel-Based Text Analysis," In Proc. of the 41st Annual Meeting of the Association for Computational Linguistics (ACL'03), pp.24-31 (2003).

服部 峻 Shun HATTORI
京都大学大学院情報学研究科社会情報学専攻博士後期課程在学中．2006 年京都大学大学院情報学研究科社会情報学専攻修士課程修了．主にモバイル・ユビキタス環境のける社会基盤技術の研究に従事．情報処理学会，電子情報通信学会，日本データベース学会各学生会員．

手塚 太郎 Taro TEZUKA
京都大学大学院情報学研究科社会情報学専攻助手. 2005年京都大学大学院情報学研究科社会情報学専攻博士後期課程修了．博士（情報学）．主に地域情報検索システム，ウェブからの知識発見，検索システムの教育への応用の研究に従事．情報処理学会，日本データベース学会各会員．

田中 克己 Katsumi TANAKA
京都大学大学院情報学研究科社会情報学専攻教授. 1976年京都大学大学院修士課程修了．博士（工学）．主にデータベース，マルチメディアコンテンツ処理の研究に従事．IEEE Computer Society，ACM，人工知能学会，日本ソフトウェア科学会，情報処理学会，日本データベース学会等各会員．