社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

カメラを用いたレイアウトフリー文書画像検索

上田　敬介†　　黄瀬　浩一†

† 大阪府立大学大学院工学研究科　〒 599–8531 大阪府堺市中区学園町 1–1
E-mail:　†ueda@m.cs.osakafu-u.ac.jp, †kise@cs.osakafu-u.ac.jp

あらまし　本稿では，我々が構築している，コンテンツ一致を基準としたカメラベースの文書画像検索手法における検索精度の向上法を提案する．我々はこれまでに，次のような特徴を持つ手法 (従来手法と呼ぶ) を提案した．すなわち，(1) 単語ごとの特徴抽出とクラスタリングを用いた単語画像の簡易コード化，(2) メッシュ特徴による特徴抽出，(3) 単語クラスタ ID の n-gram による索引付け，の 3 つである．しかし，従来手法には 3 つの問題点がある．単語の回転を考慮していないこと，フォントの変化に弱いこと，n-gram の特定性を考慮していないことである．これら対して，提案手法では回転処理を加えた単語の形状から特徴抽出を行い，更に，データベース文書を複数のフォントで登録するとともに，重み付けを加える．その結果，従来手法よりも精度の向上が見られた．レイアウトや撮影方法の異なる文書画像 320 枚をクエリ画像とし，データベースの画像 2,500 枚に対して検索実験を行ったところ，検索精度 88.1%，処理時間 671[ms] を得た．検索精度については，従来手法の 42.8%，OCR を用いた手法の 70.3%から大幅に改善しており，処理時間についても，OCR を用いた手法の 1/5 であったことから，有効性が実証された．

キーワード　文書画像検索, カメラベース, OCR, k-NN

1. はじめに

近年，電子書籍の普及により様々なサービスが提案されている．その一つとして Layered Reading [1] という情報サービスがある．これは電子書籍の紙面上に新しいレイヤを設け，付加情報を重ねて表示するというものである．しかし，印刷文書の需要は現在でも高い．そのため，現在は電子書籍に限定されているこのようなサービスを，印刷文書でも同様に享受できることが期待されている．

このようなサービスを印刷文書で可能とするためには，カメラで撮影した印刷文書から電子文書を検索する技術が必要である．実現方法の 1 つとして文書画像検索がある．文書画像検索の中には高い精度を保ちつつ実時間処理が可能なもの（例えば [2]）もあるが，概して文書のレイアウトが同一でないと検索できないという問題点がある．つまり，フォントや改行位置，行間など，レイアウトが全く同じでなければならない．これは，コンテンツが全く同じであってもレイアウトが異なるだけで，検索されないということを意味する．これらの文書を対象として検索を行うためには，上下の行などレイアウトの要素に依存しない特徴抽出をしなければならない．文書を検索する別の手法として文字認識の利用が考えられる．撮影した文書の文字を正確に認識できれば，検索に用いることができるものの，撮影画像に回転や幾何歪みなどがあった場合，文字認識をすることは容易でない．文字認識はスキャン画像を想定した手法が多く，カメラを用いた文字認識手法はまだ発展途上にある．もう一つの問題として，長い処理時間が必要にとなるという点もある．文書画像検索の目的が Layered Reading のような実時間性を要求するサービスの提供である場合，これらの問題点はより深刻になる．

上記の諸問題を解決するために，我々はコンテンツ一致の文書画像検索手法を提案した [3]．レイアウトの変動に不変な特徴を取り出すためには，文字や単語といったコンテンツの構成要素のみに依存する特徴抽出が必要となる．この要求を，文字認識のような負荷の大きい処理を施さずに満たすため，この手法では単語の簡易コード化を考える．具体的には，形状特徴を用いて単語画像をいくつかのクラスタに分類し，クラスタ ID の列として単語列をみることによって，文書画像を索引付けする．索引付けにはクラスタ ID の n-gram を用いる．ただし，この手法には，フォントの変化や単語の回転によって精度が下がるという問題点が残されている．データベース文書のフォントが 1 種類しか登録されていないため，フォントの変化にあまり強くない．また，単語の回転処理をせず，単語の輪郭のみから特徴抽出しているため，撮影画像が回転すると検索精度が下がる．そして，個々の n-gram が文書画像検索にどれほど有効かという n-gram の特定性が考慮されておらず，全ての n-gram が同等に扱われている．

本稿では，上記の問題点を解決し，この手法をより実用に近付ける．具体的には次の 3 つの改良を施す．第一の改良は，データベースの文書を複数のフォントで登録しておき，フォントの変動に頑健にすることである．第二の改良は，単語画像に回転を加え，単語の形状から特徴抽出することでロバスト性を得ることである．第三の改良は，投票時に重み付けをし，出現する n-gram の特定性を反映させることである．

本稿では，まず，関連手法の概要と問題点について説明し，

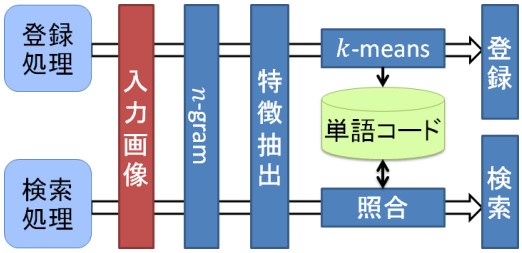


図 1　処理の流れ

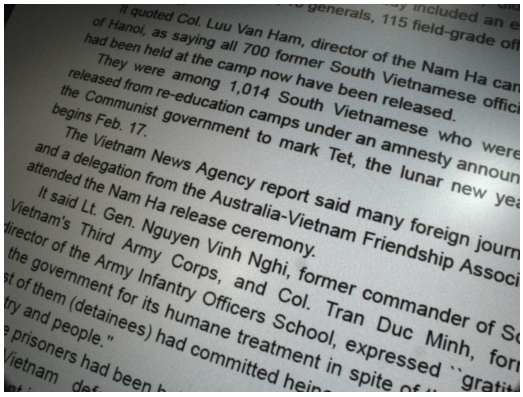


図 2　入力画像例

次に提案手法について述べる．その後，今回行った文書画像検索の実用性を検証するための実験とその結果・考察を記し，最後にまとめと今後の課題とする．

2. 従来手法

まず，提案手法の基礎となる従来手法 [3] について述べる．

2.1 方　　針

既存手法の問題点を解決するため，本研究では (1) レイアウトに依存した特徴量を使わずに検索すること，(2) 文字認識と比べて計算コストのかからない処理とすること，(3) カメラベースの入力画像にも十分対応できるロバスト性を持つことの 3 点を目標として，新しい手法を提案する．なお，将来的には多言語での動作を考えているが，現段階では第一ステップとして英文を対象とする．

まず，(1) については，次のように考える．レイアウトが変化しても改行位置を除けば単語の並びは影響を受けない．本研究ではこの点に着目し，単語の並びを特徴とする索引付けを考える．この時，(2) や (3) を実現するため，単語の種類に対して極めて少ないコードを割り当て，利用する．これにより，検索には十分ではあるものの安定した索引付けを実現する．これは，スキャナで取得した文書画像の検索に用いられる “Character Shape Code” [4] と類似の考え方である．ただし，単語を単体で用いるだけでは検索のための特定性が不足するので，単語の n-gram を考え，文書画像の索引付けに用いる．n-gram を構成する際には，連結成分の K 近傍を考慮し，可能な組み合わせを求める．これにより，撮影角度の変動にも対処を試みる．

2.2 処理の流れ

処理の流れを図 1 に示す．

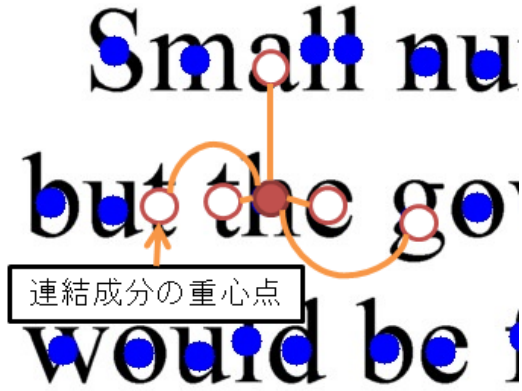


図 3　連結成分重心の K 近傍

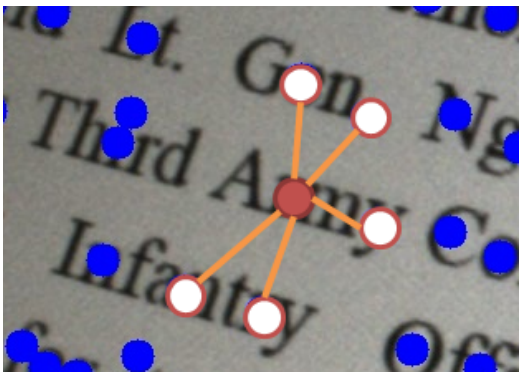


図 4　K 近傍の失敗例

最初のステップは n-gram の作成である．このためには単語の並びを求める必要がある．まず，図 2 のような文書画像を適応 2 値化した上で黒画素の連結成分を抽出し，その重心を求める．次に，この連結成分の重心を用いて K 近傍を計算することによって，ノードを連結成分，アークを K 近傍関係とするグラフを作成する．K を十分大きく取れば，文字列はこのグラフの部分グラフであると仮定できる．図 3 に例を示す．単語間の空白を考慮しても，通常，$K = 5$ 程度の近傍を考慮すれば，文字列を含めることが可能である．なお，図 4 に示すように，カメラで撮影した場合には，文字に大幅な接触があるため K の値を上げる必要がある．

次に，画像処理によって単語領域を求める．先述の処理で得た 2 値画像に対してガウス関数をたたみ込んでぼけた画像を作成し，それを再度適応 2 値化することによって，黒画素の塊を得る．これを単語領域とする．

先に求めた連結成分のグラフと単語領域に基づいて，図 5 に示す単語領域のグラフ（単語グラフ）を求める．このグラフは単語領域の重心をノード，単語領域間の近傍関係をアークとするものである．単語領域の近傍関係としては，単語中の連結成分の近傍関係を用いる．すなわち，ある単語領域に含まれる連結成分と，異なる単語領域に含まれる連結成分がアークで結ばれている時，これら 2 つの単語領域には近傍関係があるとする．

以上の単語グラフを用いると，単語 n-gram は，単語グラフの部分グラフとして求めることができる．この時，単語列が途

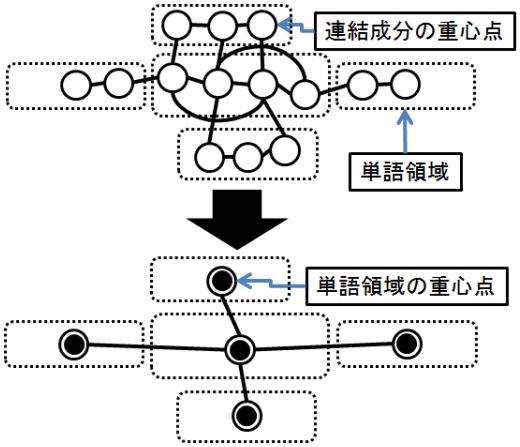


図 5　単語グラフ

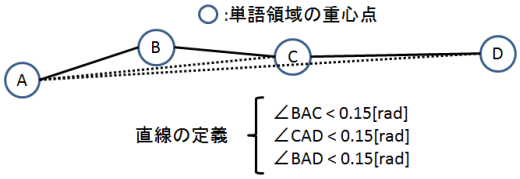


図 6　n-gram を構成するための角度の制約

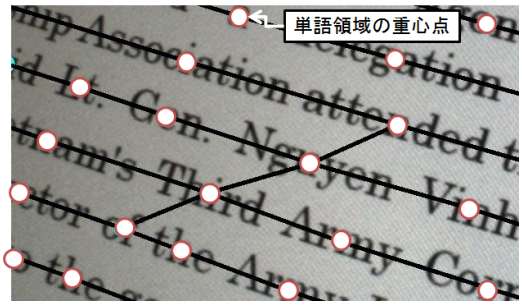


図 7　単語列の抽出

中で折れ曲がることはないと仮定し，図 6 に示すような角度の制約を満たす並びを求める．

この例に示すように，n-gram の最初の単語（図の場合は A）と他の単語の 2 つの単語で構成される角度を見て，すべてが一定の閾値以下（この場合は 0.15 radian 未満）であれば n-gram として認定する．

処理例を図 7 に示す．この図は単語 n-gram として採用されたアークを図示したものである．この例では，上記の処理により概ね単語の並びが正しく取り出されている．ただし，角度の制約によって除外された重心（この場合はカンマに相当）があるほか，“of Third Nguyen attended” のように本来の単語列とは異なる方向の組み合わせも得られることがある．単語の方向を考えるとこのような例を排除することはそれほど難しくないが，数が多くないために，従来手法では，そのまますべてを n-gram として採用する．

第 2 のステップは，単語画像コード化のための特徴抽出である．従来手法ではメッシュ特徴を用いる．

一般にカメラで撮影した画像には潰れが生じることが多い．このような場合にも安定して特徴を抽出するため，従来手法ではまず単語画像そのものではなく，先の処理で得られた単語領域の画像を特徴抽出の対象とする．この黒画素の塊を囲む矩形領域に対して，$m \times n$ のマス目を設定する．そして各マスにおいて画素値の平均を求め，それを量子化することで $m \times n$ 次元の特徴ベクトルを得る．単語画像のコード化は以下のように行う．データベース中のすべての単語画像からメッシュ特徴を取り出し，それをクラスタリングする．クラスタ数 s は実験的に定める．

検索処理の際には，上記で得られたクラスタ重心と，検索質問の単語画像から得たメッシュ特徴を照合し，検索質問の単語画像がどのコードに相当するのかを求める．

図 1 に示すように，登録処理では，単語 n-gram の生成とメッシュ特徴の抽出が終了するとそれをデータベースに登録する．n-gram は単語コードの n 個の並び $(x_1, x_2, ..., x_n)$ として表現できるので，この並びをキーとして，以下の式でハッシュ値 H_{index} を計算する．

$$H_{\text{index}} = \sum_{i=1}^{n} x_i s^{i-1} \tag{1}$$

ハッシュ表には文書 ID を登録する．ハッシュ表に登録する際に衝突が生じると，データをチェイン法で記録しておく．ただし，多くの衝突が生じることはその n-gram が十分な特定性を持たないことを意味するため，チェイン法で登録されるリストの長さが c 以上となると，そのリストを削除した上で，以後の登録を受け付けないものとする．

検索処理は次のようになる．メッシュ特徴を抽出する際に，単語の形状によっては，検索質問画像から得た単語画像が正しいクラスタに対応付かない場合が考えられる．この問題に対処するため，従来手法では，検索質問の単語画像のクラスタを 1 つに決めてしまうのではなく，メッシュ特徴が近いものから r 個の候補を用いる．すなわち，検索質問からは n 単語の各々の並びに対して，r^n 個の n-gram を生成して検索処理に用いる．

具体的な利用法は以下の通りである．各 n-gram についてハッシュ表を参照し，登録されている文書 ID を検索する．そして各文書 ID に対して 1 票ずつ投票処理を行う．最終的に最大得票となった画像を検索結果として出力する．

2.3　従来手法の問題点

従来手法の問題点として，以下の 3 点がある．

（1）　単語の回転に対処していない．

（2）　フォントの違う文書画像は正解率が下がる．

（3）　n-gram の特定性が考慮されていない．

それぞれの問題点が与える影響について説明する．

フォントによっては文字の形状が大きく異なる場合がある．フォントによる誤差を軽減するために，従来手法では単語の輪郭から特徴抽出をしている．しかし，それでは単語の識別に必要な特徴が十分得られない．また，単語の回転を考慮していないため，撮影画像が回転すると検索精度が低下してしまう．メッシュ特徴は単語の形状を量子化するため，ある程度は許容できるが，特徴抽出の方法を再考する必要がある．

文字の形状が大きく変わると，得られる特徴量が大幅に変動

する．したがって，データベースに登録している文書と全く違うフォントのクエリ画像が与えられた場合，検索精度は低下してしまう．

また，従来手法では，n-gram の出現頻度や文書中の n-gram の総数などによる特定性の違いを考慮していない．データベース文書全体で頻繁に現れるような n-gram でも数回しか現れない n-gram にも同じ 1 票を投票しているため，検索に悪影響を及ぼしている可能性がある．また，n-gram の数が少ない文書は検索しにくく，n-gram の多い文書は検索しやすいという不平等がある．

3. 提案手法

提案手法は従来手法を基にしている．上記の従来手法の問題点を解決するために，従来手法に 3 つの改良を加える．改良はそれぞれ図 1 における特徴抽出，登録，検索のステップに施す．詳細は **3.1** から **3.3** で述べる．

3.1 特徴抽出

図 8 に処理の概要を示す．

カメラで撮影した画像では単語の角度が一定ではないため，その正規化をする必要がある．まず，単語の画像に対して楕円近似を行う．その楕円の長軸方向が水平になるように単語を回転する．提案手法では，単語をぼかした画像を特徴抽出の対象とする．この単語を囲む矩形領域に対して，図 8 に示すように $m \times n$ のマス目を設定する．そして各マスにおいて画素値の平均を求め，$m \times n$ 次元の特徴ベクトルを得る．このベクトルを正規化したものをメッシュ特徴とする．

3.2 登　　録

メッシュ特徴を用いれば単語の形状を量子化することができるが，フォントによっては形状に大きな差が出る．そのため，予めデータベース文書を複数のフォントで登録しておく．ただし，データベースのメモリ量はフォントの数に比例して増えてしまう．そこで，メモリを効率よく使うために，ハッシュ関数を式 (1) から以下の式に変更する．

$$H_{\text{index}} = \left(\sum_{i=1}^{n} x_i s^{i-1} \right) \bmod H_{\text{size}} \tag{2}$$

ここで H_{size} はハッシュ表のサイズである．また，

$$\left(\sum_{i=1}^{n} x_i s^{i-1} \right) = Q H_{\text{size}} + H_{\text{index}} \tag{3}$$

となる商 Q を算出しておく．ハッシュ表には文書 ID，同一文書内での出現回数，商 Q を登録する．ハッシュ表に登録する際に衝突が生じると，データをチェイン法で記録しておく．

3.3 検　　索

各 n-gram についてハッシュ表を参照し，商 Q が一致すれば，登録されている文書 ID を検索する．そして各文書 ID に対して投票処理を行う．投票時には tf-idf 重みを適用する．従来手法では，ハッシュ値が一致すれば該当の文書 ID に 1 票投じるという単純な処理をしている．しかし，n-gram の特定性はそれぞれで異なる．同一の n-gram が特定の文書 ID から幾度も出ていれば，その n-gram はその文書を特徴づけるものであるため，特定性が高いと考えられる．よって，重み付けを重くする必要がある．また，データベース中の多数の文書に出現する n-gram は，文書の特定性が高いものではないため，あまり重要ではない．そのため，この時の 1 票は重み付けを軽くしなければならない．この 2 点を考慮した重みとして，次の式で定義される tf-idf 重みが知られている．文書 d_j における n-gram t_i の tf-idf 重み $w_{i,j}$ は以下の式で計算される．

$$w_{i,j} = \text{tf}_{i,j} \cdot \text{idf}_i \tag{4}$$

$$\text{tf}_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \tag{5}$$

$$\text{idf}_i = \log \frac{|D|}{|\{d : d \ni t_i\}|} \tag{6}$$

$w_{i,j}$ は n-gram t_i の文書 d_j における出現回数，$|D|$ は総文書数，$|\{d : d \ni t_i\}|$ は n-gram t_i を含む文書数である．

この重みを用いると，データベース中の文書 d_j は

$$\boldsymbol{d}_j = (w_{1,j}, ..., w_{m,j})$$

というベクトルで表現できる．m は出現する n-gram の種類数である．同様の表現を用いて，クエリの文書画像 q も

$$\boldsymbol{q} = (w_{1,q}, ..., w_{m,q})$$

と表すことができる．ここで，$w_{i,q}$ は

$$w_{i,q} = \begin{cases} 1 \ (n\text{-gram } t_i \text{がクエリ中に含まれている時}) \\ 0 \ (\text{それ以外}) \end{cases} \tag{7}$$

である．これらのベクトルの類似度はベクトル間の角度 θ を用いて，

$$\cos\theta = \frac{\boldsymbol{d}_j \cdot \boldsymbol{q}}{||\boldsymbol{d}_j|| ||\boldsymbol{q}||} \tag{8}$$

で表され，1 に近いほど類似していると言える．この類似度が高い文書を検索結果として出力する．

この類似度の計算は，次に示すように重み付き投票として実現できる．式 (8) は $||\boldsymbol{q}||$ が j によらないため，

$$||\boldsymbol{q}|| \cdot \cos\theta = \frac{\boldsymbol{d}_j \cdot \boldsymbol{q}}{||\boldsymbol{d}_j||} \tag{9}$$

を用いても文書の順位は変わらない．式 (9) 右辺は

$$\frac{\boldsymbol{d}_j \cdot \boldsymbol{q}}{||\boldsymbol{d}_j||} = \frac{1}{||\boldsymbol{d}_j||} \sum_{k:q_k=1} w_{k,j} \tag{10}$$

であるので，式 (10) 右辺で表される文書 d_j への重み付き投票と同じであることが分かる．

4. 実　　験

提案手法の有効性を検証するために，提案手法と従来手法，さらに OCR を用いた文書画像検索の 3 つで比較実験を行った．

4.1 実験条件

データベース文書としては AP 通信のニュース記事から 2,500

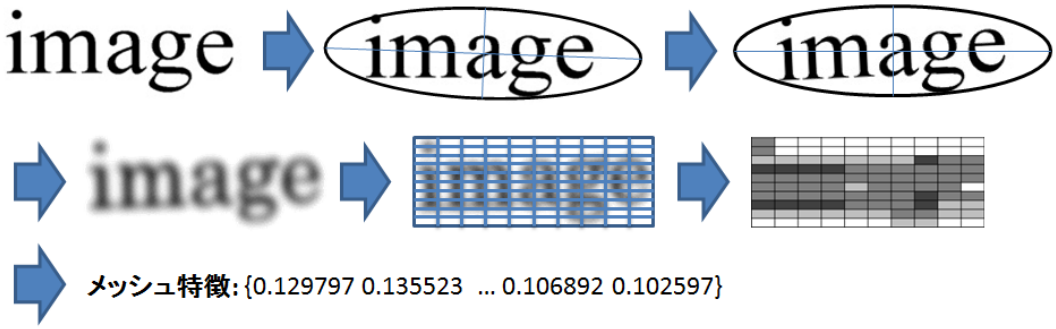
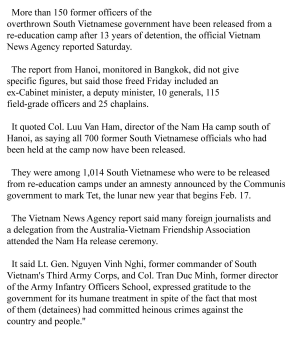


図 8　メッシュ特徴



More than 150 former officers of the overthrown South Vietnamese government have been released from a re-education camp after 13 years of detention, the official Vietnam News Agency reported Saturday.

The report from Hanoi, monitored in Bangkok, did not give specific figures, but said those freed Friday included an ex-Cabinet minister, a deputy minister, 10 generals, 115 field-grade officers and 25 chaplains.

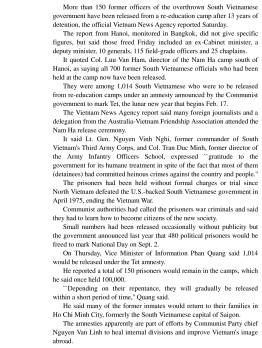
It quoted Col. Luu Van Ham, director of the Nam Ha camp south of Hanoi, as saying all 700 former South Vietnamese officials who had been held at the camp now have been released.

They were among 1,014 South Vietnamese who were to be released from re-education camps under an amnesty announced by the Communis government to mark Tet, the lunar new year that begins Feb. 17.

The Vietnam News Agency report said many foreign journalists and a delegation from the Australia-Vietnam Friendship Association attended the Nam Ha release ceremony.

It said Lt. Gen. Nguyen Vinh Nghi, former commander of South Vietnam's Third Army Corps, and Col. Tran Duc Minh, former director of the Army Infantry Officers School, expressed gratitude to the government for its humane treatment in spite of the fact that most of them (detainees) had committed heinous crimes against the country and people."

図 9　データベース文書 (一部) 例



More than 150 former officers of the overthrown South Vietnamese government have been released from a re-education camp after 13 years of detention, the official Vietnam News Agency reported Saturday.

The report from Hanoi, monitored in Bangkok, did not give specific figures, but said those freed Friday included an ex-Cabinet minister, a deputy minister, 10 generals, 115 field-grade officers and 25 chaplains.

It quoted Col. Luu Van Ham, director of the Nam Ha camp south of Hanoi, as saying all 700 former South Vietnamese officials who had been held at the camp now have been released.

They were among 1,014 South Vietnamese who were to be released from re-education camps under an amnesty announced by the Communist government to mark Tet, the lunar new year that begins Feb. 17.

The Vietnam News Agency report said many foreign journalists and a delegation from the Australia-Vietnam Friendship Association attended the Nam Ha release ceremony.

It said Lt. Gen. Nguyen Vinh Nghi, former commander of South Vietnam's Third Army Corps, and Col. Tran Duc Minh, former director of the Army Infantry Officers School, expressed ``gratitude to the government for its humane treatment in spite of the fact that most of them (detainees) had committed heinous crimes against the country and people.''

The prisoners had been held without formal charges or trial since North Vietnam defeated the U.S.-backed South Vietnamese government in April 1975, ending the Vietnam War.

Communist authorities had called the prisoners war criminals and said they had to learn how to become citizens of the new society.

Small numbers had been released occasionally without publicity but the government announced last year that 480 political prisoners would be freed to mark National Day on Sept. 2.

On Thursday, Vice Minister of Information Phan Quang said 1,014 would be released under the Tet amnesty.

He reported a total of 150 prisoners would remain in the camps, which he said once held 100,000.

``Depending on their repentance, they will gradually be released within a short period of time,'' Quang said.

He said many of the former inmates would return to their families in Ho Chi Minh City, formerly the South Vietnamese capital of Saigon.

The amnesties apparently are part of efforts by Communist Party chief Nguyen Van Linh to heal internal divisions and improve Vietnam's image abroad.

図 10　クエリ文書例

(a) Times　(b) Century　(c) Arial　(d) Verdana

図 11　各種フォント

ページを無作為に抽出して利用した．本稿では，簡単化のために複数のページをまたぐ文書はないものとする．提案手法ではデータベース文書を TimesNew Roman(以下，Times)，Century，Arial，Verdana の 4 種類のフォントで用意し，従来手法と OCR では Times のみで用意した．フォントの例をそれぞれ図 11(a)〜(d) に示す．フォントは大きく分けて “セリフ” と “サンセリフ” の 2 つに分類される．セリフとは，アルファベットをデザインする時に線の端に図 11 中の丸で囲まれた部分のような “飾り” が付いているフォントのことである．サンセリフは逆に飾りのない文字である．Times や Century はセリフ，Arial や Verdana はサンセリフである．クエリ文書にはデータベース文書のレイアウトを変更した 10 ページを作成した．クエリ文書にも同様の 4 種のフォントを用いた．データベース文書の例を図 9 に，クエリ文書の例を図 10 に示す．

次に，作成した各クエリ文書をディスプレイに表示し，正面からの撮影，仰角 15 度の撮影，右 15 度からの撮影，カメラを反時計回りに 15 度回転して撮影，さらにそれらの条件を組み合わせた $2^3 = 8$ 通り撮影した．クエリにはフォントを 4 種類用いているため，回転と射影両方がない画像と両方ある画像が 40 枚ずつ，回転と射影一方がある画像が 120 枚ずつ，計 $10 \times 4 \times 8 = 320$ 枚のクエリ画像が得られる．

表 1　最も高い検索精度とその処理時間

	検索精度 [%]	処理時間 [ms]
提案手法	88.13	671
従来手法	42.81	399
OCR	70.31	3223

提案手法の実験条件は以下の通りである．K 近傍をデータベース文書では $K = 5$，クエリ文書では $K = 8$，メッシュを 10×10，n-gram を $n = 4$ に固定した．また，単語のクラスタ数 s を $s \in \{128, 256, 512\}$ とし，検索質問の単語画像に与える単語コードの数 r を変動させ，性能を評価した．

次に，従来手法の実験条件を述べる．K 近傍と n-gram は提案手法と同値を用いる．メッシュは 3×10 で 4bit 量子化した．この時，縦を 3 分割に規定するのは，アルファベットがベースラインと x-height ラインを基に構成され，従来手法において最も安定に特徴抽出できるためである．その上で，クラスタ数 s，単語コードの数 r，ハッシュ登録数上限値 c を変動させた．

最後に，OCR の実験条件を述べる．OCR としてはオープンソースである tesseract-ocr [5] を用いた．OCR を用いた手法では，OCR により文書をコード化し，単語列を得た．そして，単語列の n-gram をすべて求め，ハッシュに格納した．こちらも，$n = 4$ に固定して実験した．この手法では提案手法と同様の重みを用いた．

評価には，検索精度と処理時間を用いた．検索精度として，クエリ全体，フォントごと，回転の有無と射影有無の組み合わせをそれぞれ比較する．処理時間としては，照合の時間だけを計測し，データベースの読み込み時間などを除外した．

4.2　結果と考察

実験の結果を表 1 から表 3 に示す．表 1 にはそれぞれ検索精度が最も高かった時の検索精度と処理時間を示す．検索精度が最も高かった時のパラメータは，提案手法では $s = 256$，$r = 5$ の時，従来手法では $s = 256$，$r = 5$，$c = 1$ の時であった．表 2 と表 3 は検索精度が最も高かった時の詳細な結果である．表 2 にはフォントごとの検索精度を示す．表 3 は回転の有無と射

表 2　フォントごとの検索精度 [%]

	フォント			
	セリフ体		サンセリフ体	
	Times	Century	Arial	Verdana
提案手法	83.75	95.00	85.00	88.75
従来手法	46.25	48.75	36.25	40.00
OCR	65.00	70.00	68.75	77.50

表 3　回転と射影による検索精度 [%]

	回転なし 射影なし	回転なし 射影あり	回転あり 射影なし	回転あり 射影あり
提案手法	97.50	90.00	82.50	85.00
従来手法	87.50	84.17	0.00	0.83
OCR	100.00	98.33	70.00	32.50

影の有無に分けた検索精度である．

まず，提案手法と従来手法を比較する．表 1 と表 2 より，従来手法は処理速度は速いものの，検索精度が低く，フォントによって精度に少し偏りがあることが分かる．表 3 を見ると，従来手法は回転がなければ射影の有無に関わらず，高い検索精度が得られている．これは，文書画像全体で大きく射影がかかっていても，一単語ごとに見ればあまり大きな影響はないためと考えられる．しかし，回転に対しては全く検索できていない．これは，回転によってメッシュ特徴が大きく変動したためと思われる．一方，提案手法は，回転に対する処理を施したことにより，従来手法よりは処理時間が増えているが，回転のある画像に対しても検索精度を維持できている．加えて，提案手法では，回転と射影両方のある画像に対しても精度があまり低下していない．これは，提案手法の楕円近似による正規化が回転や射影による幾何学的変動に対して頑健であることを意味している．また，提案手法は従来手法に比べて，フォントによる偏りが小さくなった．つまり，データベース文書のフォントを予め複数用意しておくことにより，フォントの変動に対して頑健な処理が実現可能であると分かった．従来手法ではハッシュ衝突回数上限値が $c = 1$ の時最も高い精度が得られた．これは，検索に有用ではない特定性の低い n-gram が多く存在することを意味するが，中には同一の文書から複数回出現するような特定性の高い n-gram も除去している可能性がある．そのため，同一の n-gram でも重みを変えて検索に用いることが重要である．

次に，提案手法と OCR を比較する．OCR は，本来スキャン画像を想定した技術であるが，表 1 を見ると，撮影画像に対しても一定の精度が得られていることが分かる．しかし，提案手法の約 5 倍の処理時間を要している点が問題である．表 3 より，回転のない画像に対しては提案手法よりも高い精度が得られている．しかし，回転のある画像では精度が低下し，更に射影も加わると精度が大きく下がる．この回転と射影のある画像の詳細を見ると，下から撮影した画像は 67.5%であるのに対し，横から撮影した画像は 17.5%，斜め下から撮影した画像は 12.5%という精度であった．OCR 処理では，ベースラインやアセンダライン (小文字の f，h，l などの上端に引かれる水平線)，x-height ライン，ディセンダライン (小文字の g，p，y などの下端に引かれる水平線) などのラインを推定することが，正しく文字を認識する上で重要である．このラインの推定には，全てのラインが互いに平行であるという仮定が用いられている．ところが，幾何歪みのある画像では，もはやこの仮定が成り立たない．したがって，ラインの推定に失敗し，文字認識が困難になる．その結果，検索精度が低下していると思われる．また，射影がかかると，文書画像では文字の接触が起きやすくなる．OCR は文字が接触すると正確に文字認識することが難しくなるため，これも検索精度が落ちる原因である．

以上の結果より，提案手法は従来手法と比較して，フォントによる検索精度の偏りを軽減し，回転に対するロバスト性を実現し，検索精度を向上させることができた．また，比較手法として挙げた OCR と比較すると，1/5 の処理時間で OCR よりも高い検索精度が得られた．OCR を用いるよりも高速ではあるが，リアルタイムで用いるためには，更なる処理時間の削減が必要である．また，データベースをより大規模にすることも不可欠である．

5. ま と め

我々が構築しているコンテンツ一致の基準による文書画像検索手法は，単語クラスタの n-gram に基づく検索法を用いている．しかし，我々の従来手法では，単語の回転に対処していない，フォントの異なる文書に対して検索精度が落ちる，n-gram の特定性が考慮されていないという問題点があった．本稿では，従来手法に単語に回転処理を加えること，予め複数のフォントでデータベースを作成しておくこと，n-gram に重み付けを施すことの 3 点を導入した．その結果，処理時間 671[ms] で 88.1%の検索精度を得た．これは，従来手法の 42.8%，OCR を用いた手法のの 70.3%から改善しており，また，OCR を用いた場合の 1/5 の処理時間であったことから，提案手法の有効性が検証された．今後の課題としては，データベースの大規模化，処理時間の削減が挙げられる．

謝　　辞

本研究の一部は日本学術振興会科学研究費補助金基盤研究 (B)(22300062)，ならびに JST CREST の補助による．

文　　献

[1] http://84dialog.blogspot.com/2010/03/layered-reading.html, 2011

[2] Kazutaka Takeda and koichi Kise and Masakazu Iwamura, “Real-Time Document Image Retrieval for a 10 Million Pages Database with a Memory Efficient and Stability Improved LLAH,” Proceedings of the Int’l Conf. on Document Analysis and Recognition, pp.1054-1058, Sep. 2011.

[3] 上田 敬介, 黄瀬 浩一, “レイアウト変動にも対応できる文書画像検索法”, 電子情報通信学会技術研究報告, vol.111, PRMU2011-111, pp.25–30, Nov.2011.

[4] H. Bunke and P. S. P. Wang Eds.: “Handbook of Character Recognition and Document Image Analysis”, World Scientific Pub Co Inc (1997).

[5] http://code.google.com/p/tesseract-ocr/