色とその配置位置に基づいた視覚的印象による Webページ検索手法の実現方式

石川　幹直†　　細川　宜秀†　　高橋　直久†

† 名古屋工業大学大学院 工学研究科 情報工学専攻 〒 466-8555 愛知県名古屋市昭和区御器所町
E-mail: †ishikawa@moss.elcom.nitech.ac.jp, ††{hosokawa,naohisa}@nitech.ac.jp

あらまし　本研究では，Web ページに含まれる色とその配置位置に基づいた視覚的印象による Web ページ検索手法を提案する．ここで，色とその配置位置に基づいた視覚的印象とは，Web ページがブラウザに表示された直後のイメージのうち，検索者が記憶する部分を指す．本手法の主要な特徴は，Web ページ画像を視覚的印象として，Web ページ検索のための索引自動構成機構を実現する点にある．これにより，検索者は色とその配置位置を記憶している状況下において，検索キーを言葉で与えることが困難な場合でも Web ページ検索が可能となる．さらに，Web ページ画像から抽出される特徴を検索者が指定する機構を実現する．これにより，同一 Web ページに対する異なる検索者の記憶の相違による検索精度の低下の抑制が可能になる．本手法のプロトタイプを実装し，実験により本手法の有効性を検証する．

キーワード　情報検索，Web 利用技術，画像処理

An Implementation Method of the Web Page Retrieval based on the Visual Impression

Motonao ISHIKAWA†, Yoshihide HOSOKAWA†, and Naohisa TAKAHASHI†

† Department of Computer Science and Engineering Graduate School of Engineering,
Nagoya Institute of Technology
Gokiso, Showa, Nagoya, 466-8555 Japan
E-mail: †ishikawa@moss.elcom.nitech.ac.jp, ††{hosokawa,naohisa}@nitech.ac.jp

Abstract　In this paper, we present an implementation method of the Web page retrieval based on the visual impression. This method equips us search the Web page using color scheme of that one as a search key. Our main feature is to conform visual effects of a Web page to that of a indistinct memory. And we implement our system to make pictures when we input indistinct memory of color scheme of the Web page. By using our system, we can search the Web page using the visual impression instead of the keywords. We clarify the validity of our method by several experiments.

Key words　Information Retrieval, Web Utilization, Image Processing

1. はじめに

近年，文書データだけでなく，画像やそれらを配置するレイアウトで構成された Web ページが増加している．現在までに，Web ページが文書データであることを利用し，キーワードによる検索で Web ページの獲得が可能となった．しかしながら，Web ページをキーワードで表現することが困難な状況が存在する．これは，Web ページはテキストデータ，色，画像などのコンテンツと，それらを配置するレイアウト情報を特徴に持つマルチメディアデータであることに起因する．そのため，Web ページのテキストデータに含まれる内容を記憶しておらず，他の特徴を記憶している状態においては，記憶に残っている Web ページの状態をキーとして Web ページを再現するのが自然な要求である．したがって，キーワードによる検索だけではなく，画像，レイアウト，見た目に応じた検索機能の実現は，検索者が欲する Web ページへのアクセス機会を増大させる．

本章では，Web ページに含まれる色とその配置位置に基づいた視覚的印象による Web ページ検索手法を提案する．ここで，色とその配置位置に基づいた視覚的印象とは，Web ページがブラウザに表示された直後のイメージ（以後 Web ページ画像と

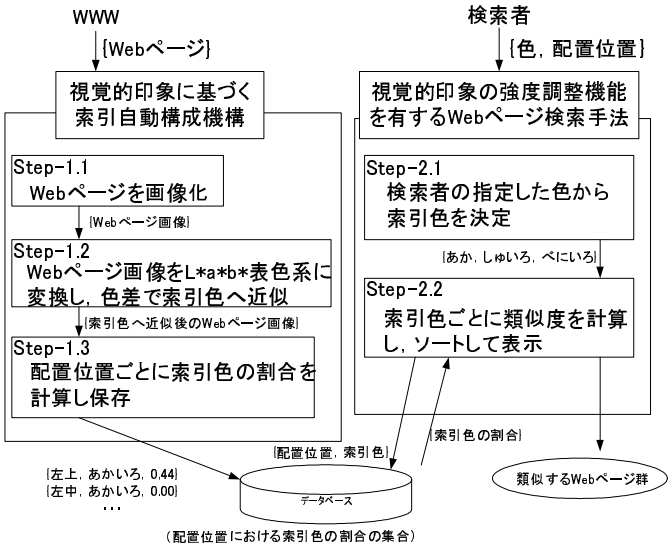

図 1　提案手法の全体構成

呼ぶ）のうち，検索者が記憶する部分を指す．本手法の主要な特徴は，Web ページ画像を視覚的印象として，Web ページ検索のための索引自動構成機構を実現する点にある．これにより，検索者は色とその配置位置を記憶している場合に，検索キーを言葉で与えることが困難な状況において Web ページ検索が可能となる．さらに，Web ページ画像から抽出される特徴を検索者が指定する機構を実現する．これにより，同一 Web ページに対する異なる検索者の記憶の相違による検索精度の低下の抑制が可能になる．本手法のプロトタイプを実装し，実験により本手法の有効性を検証する．

2. 提案手法の実現方式

Web ページ画像は，画像検索を目的としたシステムで対象となる写真群と比較すると次の 2 点の特徴を持つ．

特徴-1　モニタへ色を再現させる制限を持つため，単調な色が連続配置している．

特徴-2　左部にメニューボタンを並べる，上部に企業の色を入れる，などある程度規則性のあるレイアウトを持つ．

提案システムは，Web ページ画像が持つ上記 2 点の特徴を視覚的印象として抽出し，その特徴に基づいた Web ページ検索のための索引自動構成機構（視覚的印象に基づく索引自動構成機構と呼ぶ）を実現する．視覚的印象の抽出は，上記特徴に従い “色とその配置位置” であると仮定して行う．さらに，検索者毎に異なる記憶に対応した Web ページ検索手法（以後視覚的印象の強度調整機能を有する Web ページ検索機構と呼ぶ）を提供する．

2.1　視覚的印象に基づく索引自動構成機構の実現方式

ここでは，Web ページ画像の特徴的部分を視覚的印象として抽出し，その特徴的部分に基づいた Web ページ検索のための索引自動構成機構の実現方式について述べる．図 1 の左部分に，索引自動構成機構の構成を示す．本手法は，次の手順により実行する．

Step1-1　Web ページ画像を生成する

閲覧した Web ページを画像として保存する．検索者が記憶している部分は実際に表示された部分であることから，ブラウザで表示された部分のみを画像として保存する．閲覧時における Web ページの情報をそのまま記憶することで，検索者が実際に検索を行う時の環境と，閲覧時の環境は同等であると考えられ，閲覧時と検索時の印象の違いは少なくなる．

Step1-2　Step1-1 で得られた Web ページ画像を索引となる色へ近似する

提案手法は，Web ページ画像に含まれる色の割合を索引とする（以後，索引となる色を索引色と呼ぶ）．Web ページ画像に含まれる色から索引色への変換は，1 ピクセルごとに含まれる RGB 値を L*a*b*表色系へ変換し，索引色との色差がもっとも小さいものへ近似することで行う．L*a*b*表色系は，RGB 値からの変換式が用意されており，色差が一定の空間であるので色の違いが距離により比較可能となる利点がある．以下の変換式を用いて RGB 値を L*a*b*値へ変換する [1], [9].

$$X = 0.3933R + 0.3651G + 0.1903B \tag{1}$$

$$Y = 0.2123R + 0.7010G + 0.0858B \tag{2}$$

$$Z = 0.0182R + 0.1117G + 0.9570B \tag{3}$$

$$L* = 116\left(\frac{Y}{Y_n}\right)^{\frac{1}{3}} - 16 \tag{4}$$

$$a* = 500\left[\left(\frac{X}{X_n}\right)^{\frac{1}{3}} - \left(\frac{Y}{Y_n}\right)^{\frac{1}{3}}\right] \tag{5}$$

$$b* = 200\left[\left(\frac{Y}{Y_n}\right)^{\frac{1}{3}} - \left(\frac{Z}{Z_n}\right)^{\frac{1}{3}}\right] \tag{6}$$

ここで式 (4),(5),(6) における X_n, Y_n, Z_n は Web ページ閲覧時の環境により異なる刺激値を補正する定数であり，ここでは文献 [9] に従い $X_n = 98.072, Y_n = 100, Z_n = 118.225$ とした．ただし式 (4)，(5)，(6) における $(X/X_n)^{1/3}, (Y/Y_n)^{1/3}, (Z/Z_n)^{1/3}$ は，0.2069 以下の場合にそれぞれ以下の式に置き換わる．

$$\left(\frac{X}{X_n}\right)^{\frac{1}{3}} \rightarrow 7.787\left(\frac{X}{X_n}\right) + \frac{16}{116}$$

$$\left(\frac{Y}{Y_n}\right)^{\frac{1}{3}} \rightarrow 7.787\left(\frac{Y}{Y_n}\right) + \frac{16}{116}$$

$$\left(\frac{Z}{Z_n}\right)^{\frac{1}{3}} \rightarrow 7.787\left(\frac{Z}{Z_n}\right) + \frac{16}{116}$$

任意の 2 色 $c_1(L_1, a_1, b_1)$,$c_2(L_2, a_2, b_2)$ の色差 $\Delta\acute{E}(c_1, c_2)$ は式 (7) で計算可能である．

$$\Delta\acute{E}(c_1, c_2) = \sqrt{(L_1 - L_2)^2 + (a_1 - a_2)^2 + (b_1 - b_2)^2} \tag{7}$$

式 (1)–(6) を用いて，Web ページ画像に含まれるすべての画素の RGB 値を L*a*b*表色系に変換する．その後式 (7) により索引色との色差を計算し，Web ページ画像に含まれる色を，索引色の中で色差が最も小さいものに近似する．提案手法では，配置位置に含まれる索引色に対し検索者が選択した索引色との色差を用いて重み付けを行う．ある索引色 c_1 が選択された時 c_2 につける重み $\Delta E(c_1, c_2)$ を式 (8) と定義する．

$$\Delta E(c_1, c_2) = \frac{\Delta\acute{E}(c_1, c_2)}{(\text{索引色間距離の最大値})} \tag{8}$$

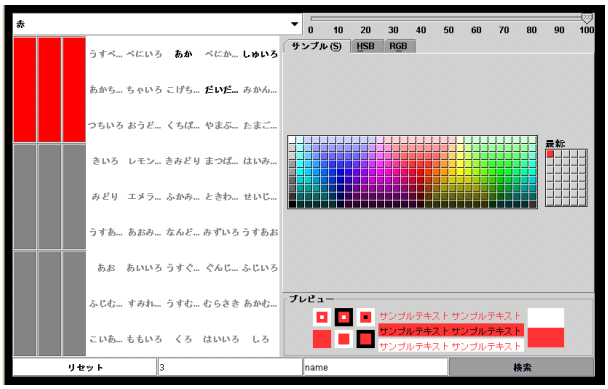

図 2　検索者が視覚的印象を入力するインタフェース

Step1-3　Step1-2 で得られた Web ページ画像の色の配置位置ごとにおける割合を計算し，索引として保存する

提案手法では，検索者は視覚的印象を元に色と，その色の配置位置を入力することで Web ページ検索を可能としている．Web ページ画像を上下左右 3 等分し，合計 9 箇所を色の配置位置とする．各配置位置ごとに含まれる索引色の割合を元に類似度を計算し，Web ページの索引とする．すなわち，1 つの Web ページに対し配置位置 P_i に対する索引色 C_{ij} を生成する．ただし P_i は Web ページ画像を分割後における i 番目 $(1 \leqq i \leqq (\text{索引色数}))$ の配置位置，C_{ij} は配置位置 P_i に対し j 番目 $(1 \leqq j \leqq 9)$ の索引色が含まれる割合である．

2.2　視覚的印象の強度調整機能を有する Web ページ検索手法の実現方式

ここでは，視覚的印象に基づく索引自動構成機構により索引をつけられた Web ページから，検索者毎に異なる記憶に対応した Web ページ検索手法について述べる．図 1 右部分に，本手法の構成を示す．本手法は，次の手順により実行する．

Step2-1　検索者が与えた色，配置位置を元に索引色を決定する

検索者が視覚的印象を入力するインタフェースを図 2 に示す．検索者は，以下の手順で視覚的印象をシステムに入力する．

視覚的印象入力手順-1　検索したい色を選択する

図 2 左上部のリストから，検索したい色を選択する（図 2 では “赤” を選択している）．ここで選択可能な色は，{ 赤, ピンク, 青, 黄, 緑, オレンジ, 茶, 黒, 白 } とした．

視覚的印象入力手順-2　手順-1 で指定した色を定義する

手順-1 で指定した色を，図 2 右部の色選択部からすべて選択する．選択した色と，索引色との色差が小さいものを，検索者が選択した索引色とする（図 2 ではあかいろ，しゅいろ，だいだいいろが索引色として定義された.）．この手順により，検索者の環境による色の表示の差異をなくし，検索者の記憶から適切な索引色を選択させることを可能とする．

視覚的印象入力手順-3　色の配置位置を指定する

図 2 左部の格子が Web ページに対応する．手順-1，手順-2 で指定した色が Web ページ上のどこに存在していたかを選択する（図 2 では左上，中上，右上を選択した.）．

Step2-2　検索者が与えた色の選択した配置位置における割合，選択していない配置位置における割合を元に類似度を計算し，ソートして表示する

図 3　検索者の視覚的印象を用いた Web ページ検索結果

検索者が入力した色とその配置位置から，Web ページの類似度を決定する．類似度 Sim は，検索者が指定した配置位置に含まれる各索引色の割合を元に構成される α（以後 α 類似度と呼ぶ）と，検索者が指定していない配置位置における各索引色の割合を元に構成される β（以後 β 類似度と呼ぶ.）を用いて式 (9) で定義する．

$$Sim = \alpha + k \cdot \alpha \cdot f(\beta) \tag{9}$$

ただし k は感性定数であり，$0 \leqq k \leqq 10$ の値を取る．検索者が指定するパラメータである． α 類似度は検索者が記憶している部分に対応し，β 類似度は検索者が記憶していない部分に対応する． ここで α 類似度と β 類似度を用いるのは，以下の仮説によるものである．

仮説-1　検索者が選択していない配置位置における索引色は，検索者が選択した配置位置における索引色を引き立たせる．

また，検索者が指定した索引色と，配置位置に含まれる全索引色との色差を用いて重み付けを行う．我々は，色差に対して以下の仮説を立てた．

仮説-2.1　検索者が指定した配置位置における選択索引色との色差が小さい索引色ほど強く印象に残る．

仮説-2.2　検索者が指定していない配置位置における選択索引色との色差が大きい索引色ほど選択索引色を引き立たせる．

これら仮説に基づき，検索者が k 番目の索引色を選択した場

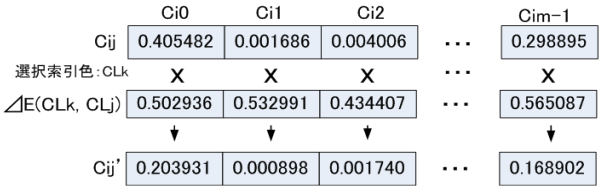

図 4　色差で重み付けを行う β 類似度の算出プロセス．ただし $C_{ij}, CL_i, \Delta E(CL_i, CL_j), C'_{ij}$ はそれぞれ Web ページ画像の i 番目の配置位置における j 番目の索引色が含まれる割合，i 番目の索引色の L*a*b*値，最長索引色間距離に対する 2 索引色間の割合，距離による補正後の索引色の割合を表す．ここで得られる C'_{ij} の総和を β 類似度とする．

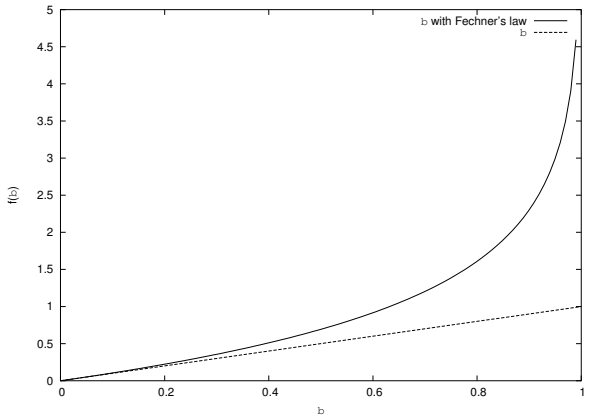

図 5　β 類似度と感覚の強さの関係．β 類似度は検索者が選択していない配置位置に含まれる，選択索引色以外の索引色で構成される．β 類似度が大きいほど式 (14) に従い選択索引色を強調させる．

合，i 番目の配置位置における α 類似度 α_{ik}，β 類似度 β_{ik} は次の式で定義する．

$$\alpha_{ik} = \sum_{j=0}^{allC-1} C_{ij} \cdot (1 - \Delta E(CL_k, CL_j)) \tag{10}$$

$$\beta_{ik} = \sum_{j=0}^{allC-1} C_{ij} \cdot \Delta E\,(CL_k, CL_j) \tag{11}$$

ただし $allC, CL_k$ はそれぞれ総索引色数，k 番目の索引色の L*a*b*値を表す．図 4 に β 類似度算出のプロセスの例を示す．式 (10),(11) を Web ページ画像全体に適用した α,β は次の式で定義する．

$$\alpha = \sum_{i=0}^{allP-1} \sum_{k=0}^{allC-1} \alpha_{ik} \cdot ChC_k \cdot ChP_i \tag{12}$$

$$\beta = \sum_{i=0}^{allP-1} \sum_{k=0}^{allC-1} \beta_{ik} \cdot ChC_k \cdot (1 - ChP_i) \tag{13}$$

ただし $allP$ は Web ページ画像の分割数である．また，ChC_k, ChP_i は検索者が選択した索引色，配置位置であるかどうかを表し，選択していれば 1，選択していなければ 0 とする．

式 (9) における $f(\beta)$ は，“感覚は刺激の対数に比例する” というフェヒナーの法則に従ったもので，次の式で定義する．

$$f(\beta) = -\log(1 - \beta) \tag{14}$$

図 5 に β 類似度として式 (14) を用いた場合と β をそのまま用いた場合との違いを示す．β 類似度は検索者が選択していない配置位置における，選択索引色以外の色の割合を元に構成されている．β 類似度が大きいほど強い刺激が与えられることになる．

式 (9) を視覚的印象に基づく自動索引構成機構を用いて生成された全索引に対して適用し，類似度 Sim の大きい順にソートして表示する．

3. 実　　験

提案手法は，検索者が選択した配置位置における各索引色の割合を元に構成される α 類似度と，検索者が指定していない配置位置における各索引色の割合を元に構成される β 類似度を用いて，検索者がインタフェースにより入力した視覚的印象と，登録された Web ページとの類似度を計算する．本手法では，Web ページの持つ視覚的印象について次の 3 つの仮説を用いている．

仮説-1　検索者が選択していない配置位置における索引色は，検索者が選択した配置位置における索引色を引き立たせる．

仮説-2.1　検索者が指定した配置位置における選択索引色との色差が小さい索引色ほど強く印象に残る．

仮説-2.2　検索者が指定していない配置位置における選択索引色との色差が大きい索引色ほど選択索引色を引き立たせる．

実験では，次の 4 つの項目を検証することで，これら仮説を検証し，提案手法の妥当性を検証する．

検証-1　β 類似度導入による提案手法の優位性検証

本検証では，検索者が指定していない配置位置における選択索引色が，検索者が指定した配置位置における選択索引色を引き立たせることを示し，提案手法の優位性を検証する．

検証-2　感性定数 k の強弱による提案手法の妥当性検証

本検証では，β 類似度の強弱により，特徴的な色を選択した配置位置に引き立たせることを示し，提案手法の妥当性を検証する．

検証-3　色同士の距離を用いた重み付けの効果検証

本検証では，選択索引色との色差を用いた重み付けの効果を検証する．

検証-4　検索キーの違いによる提案手法の精度検証

本検証では，検索キーとして用いる色の違いによる検索精度の違いを検証する．

3.1　実 験 方 法

実験では，各検証のために以下の 3 手法に従い，提案システムを構成する．

提案手法-1　式 (9) に基づく類似度計算機構を備えた視覚的印象による Web ページ検索手法

提案手法-1 では，式 (9) に基づいて類似度を計算する．これにより α 類似度，β 類似度ともに色差を含めた類似度計算を行う．

提案手法-2　β のみに色差の重み付けを行う類似度計算機構を備えた視覚的印象による Web ページ検索手法

提案手法-2 では，式 (12) における α_{ik} は式 (10) の代わりに

以下の式を用いる．

$$\alpha_{ik} = C_{ij} \tag{15}$$

式 (15) により，α 類似度は，検索者が指定した配置位置に含まれる選択索引色の割合で表される．β 類似度は提案手法-1 と同様である．

提案手法-3　色差の重み付けを行わない類似度計算機構を備えた視覚的印象による Web ページ検索システム

提案手法-3 では，色差による重み付けを行わない．式 (12) における α_{ik} には式 (15) を用いる．式 (13) における β_{ik} には式 (11) の代わりに以下の式を用いる．

$$\beta_{ik} = 1 - C_{ij} \tag{16}$$

これは，検索者が指定していない配置位置の色は，検索者が指定した色以外の色が均等に色の印象に影響を与えることを意味する．式 (16) により，β 類似度は，検索者が指定していない配置位置に含まれる，選択索引色と異なる索引色の割合の総和で表される．

被験者は，上記 3 手法を備えた Web ページ検索システムを用いて，色とその配置位置を入力することにより Web ページを検索する．その後，システムが出力した結果から被験者の要求した色と配置位置に適した Web ページを選択して正解とする．これにより，上位 n 件の適合率を求めて，提案手法の優位性，効果の検証に用いる．ここで適合率は以下の式で定義する．

$$（適合率）= \frac{（上位\ n\ 件で正解の\ \mathrm{Web}\ ページ数）}{n} \tag{17}$$

提案手法は，索引色となる色の種類を，JIS 規格で定められている色鉛筆が持つ有彩色 42 色 [8] と，白，黒，灰の 45 色で構成した．実験に用いる Web ページには，文書データ，画像，それらを構成するレイアウトを持った Web ページをランダムに 1200 件収集したものを用いた．

実験には 7 人の大学生が被験者として参加した．各被験者は，“上部が赤い Web ページ”，“上部が青い Web ページ”，“上部が黄色い Web ページ”，“上部が緑の Web ページ” を要求することとした．

各検証のため，式 (9) における感性定数 k がそれぞれ=0,1,10 の場合において，3 種類の提案手法を用いて検索を行い，適合率を求めた．k=0 は，β 類似度を用いない場合に相当する．また，k=10 は β 類似度を強調した場合に相当する．

検証-1 は，k=0 の結果と k=1 の結果を比較することにより，β 類似度導入による効果を検証する．検証-2 は，k=1 の結果と k=10 の結果を比較することにより，β 類似度強調による効果を検証する．検証-3 は，提案手法毎の類似度を比較することにより，適合率の変化を検証する．

3.2　実験結果と考察

表 1 は，被験者 7 人が Web ページ検索を行った時の適合率の平均を表す．各提案手法において，検索結果の例を図 7–図 10 に示す．

3.2.1　β 類似度導入による提案手法の優位性に対する考察

表 1 における，k=0 の結果と k=1 の結果を比較することにより，β 類似度導入による効果を検証する．図 7–図 10 より，k=0 の場合には検索者が選択した配置位置以外において，選択索引色の含まれる割合が多い Web ページが上位にあがるのに対し，k=1 の時は，検索者が選択した配置位置における色のみが検索者の指定した色となる Web ページが上位にランキングされた．また，表 1 において k=1，すなわち β 類似度の導入によって適合率が向上した．これにより，β 類似度を導入することによって，Web ページ検索の精度が向上することを明らかにした．

表 1　k の値による適合率の変化

(k = 10)	提案手法-1	提案手法-2	提案手法-3
上位 10 件	44.71 %	48.24 %	58.82 %
上位 20 件	42.94 %	48.24 %	53.23 %
上位 30 件	40.78 %	48.24 %	52.55 %
(k = 1)	提案手法-1	提案手法-2	提案手法-3
上位 10 件	34.21 %	41.33 %	38.24 %
上位 20 件	31.58 %	44.33 %	45.00 %
上位 30 件	30.00 %	46.00 %	48.24 %
(k = 0)	提案手法-1	提案手法-2	提案手法-3
上位 10 件	21.25 %	28.75 %	26.88 %
上位 20 件	24.69 %	35.63 %	35.63 %
上位 30 件	21.88 %	38.13 %	38.13 %

図 6　図 8 における α 類似度への色差を用いた重み付け導入による 2 位の変化 (k=10)．左が α 類似度に重み付けを行わない結果，右が α 類似度に重み付けを行った結果．検索者が記憶している色が “青” なのに対し，近似された “水色” が上位にランキングされた．被験者は “水色” をイメージしていなかったため，不適合とみなした．

3.2.2　感性定数 k の強弱による提案手法の妥当性に対する考察

表 1 における k=1 の結果と k=10 の結果を比較することにより，β 類似度を強調することによる効果を検証する．表 1 より，いずれの提案手法においても，k=1 の時と比較して k=10 の時の適合率が向上した．また，図 7–図 10 より，k の値を上げることによって，検索者が指定した配置位置に，検索者が指定した色を特徴に持つ Web ページが上位にランキングされることを確認した．これにより，提案手法の妥当性を確認した．

3.2.3　色同士の距離を用いた重み付けの効果に対する考察

表 1 における提案手法毎の類似度を比較することにより，色同士の距離を用いた重み付けの効果を検証する．表 1 より，提案手法-1 は，他の 2 手法と比較すると適合率が低下している．これは，α 類似度の算出で用いる検索者の選択した配置位置における選択索引色が，図 2 を介して正確にシステムに入力された色であり，この色を近似することで検索者が指定した色が近

図 7　k の値によるシステムの出力結果の変化（上が赤の Web ページを検索）

表 2　赤色をキーとした Web ページ検索時の適合率

(k = 10)	提案手法-1	提案手法-2	提案手法-3
上位 10 件	60.00%	50.00%	45.00%
上位 20 件	52.55%	61.25%	55.00%
上位 30 件	40.83%	57.50%	58.33%
(k = 1)	提案手法-1	提案手法-2	提案手法-3
上位 10 件	37.50%	45.00%	32.50%
上位 20 件	38.75%	58.75%	51.25%
上位 30 件	34.17%	59.17%	53.33%
(k = 0)	提案手法-1	提案手法-2	提案手法-3
上位 10 件	25.00%	37.50%	42.50%
上位 20 件	35.00%	47.50%	46.25%
上位 30 件	30.00%	48.33%	49.17%

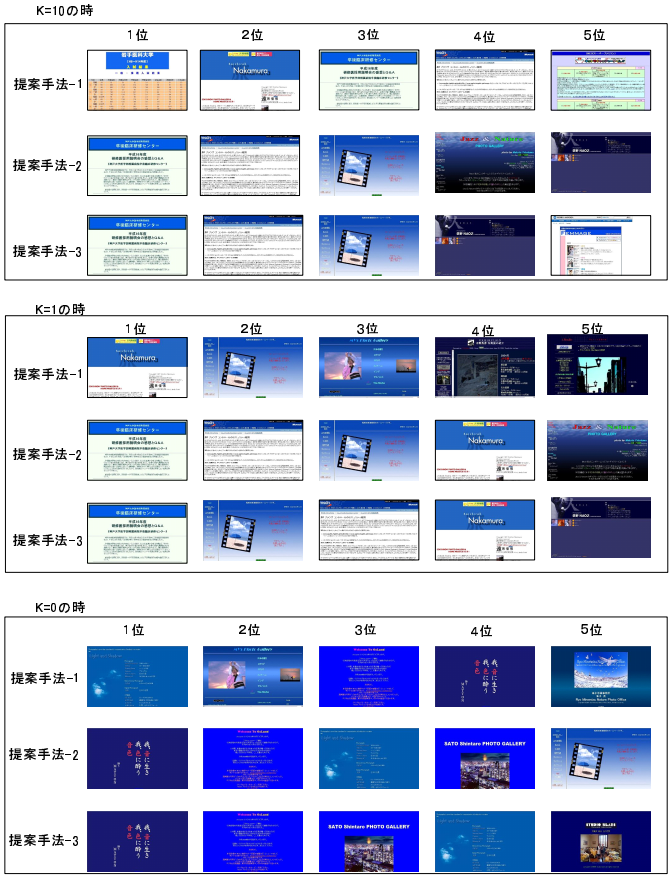

図 8　k の値によるシステムの出力結果の変化（上が青の Web ページを検索）

表 3　青色をキーとした Web ページ検索時の適合率

(k = 10)	提案手法-1	提案手法-2	提案手法-3
上位 10 件	78.00%	73.33%	80.00%
上位 20 件	79.00%	66.67%	64.00%
上位 30 件	78.67%	58.89%	54.00%
(k = 1)	提案手法-1	提案手法-2	提案手法-3
上位 10 件	45.71%	63.33%	52.00%
上位 20 件	40.71%	50.00%	43.00%
上位 30 件	39.52%	42.22%	42.00%
(k = 0)	提案手法-1	提案手法-2	提案手法-3
上位 10 件	27.50%	30.00%	27.50%
上位 20 件	30.00%	31.25%	31.25%
上位 30 件	26.67%	29.17%	33.33%

似され，検索者が記憶している色と，提案手法-1 の出力する色とに差異が生じてしまうことが原因と考えられる．例として図 6 を取り上げる．被験者は “青” を検索しようとした時，α 類似度に色差の重みづけを行わない結果表示された図 6 左が，図 6 右に変化した．図 6 右は色が検索者の指定した “青” とは異なる色であるものの，“青” と類似した色が選択した配置位置に多く含まれているため，上位にランキングされて適合率を低下させた．このことから，提案手法は検索者が検索キーとしてイメージする色を正確にシステムに伝えており，α 類似度に色差の重み付けを行わない方が精度が向上することを明らかにした．

また，提案手法-2 による適合率と，提案手法-3 による適合率にはほとんど差がない．このことから，β 類似度に色差を用いた重み付けを行う必要がないことを明らかにした．

3.2.4　検索キーの違いによる提案手法の精度に対する考察

表 2–表 5 に，検索に用いた色ごとの適合率の平均を示す．検索キーに用いる色の種類によって，適合率にばらつきが生じた．図 7–図 10 より，検索者が指定したキーと類似した索引色が選択されながら，類似した索引色と検索者との相違が生じているために適合率が低下していることが明らかとなった．図 8 では，“青” の類似した索引色として “すみれ”“ぐんじょう” を多く含む Web ページが上位にランキングされ，検索者はそれを青に適合としている．しかしながら，図 9 では，“黄” に類似した索

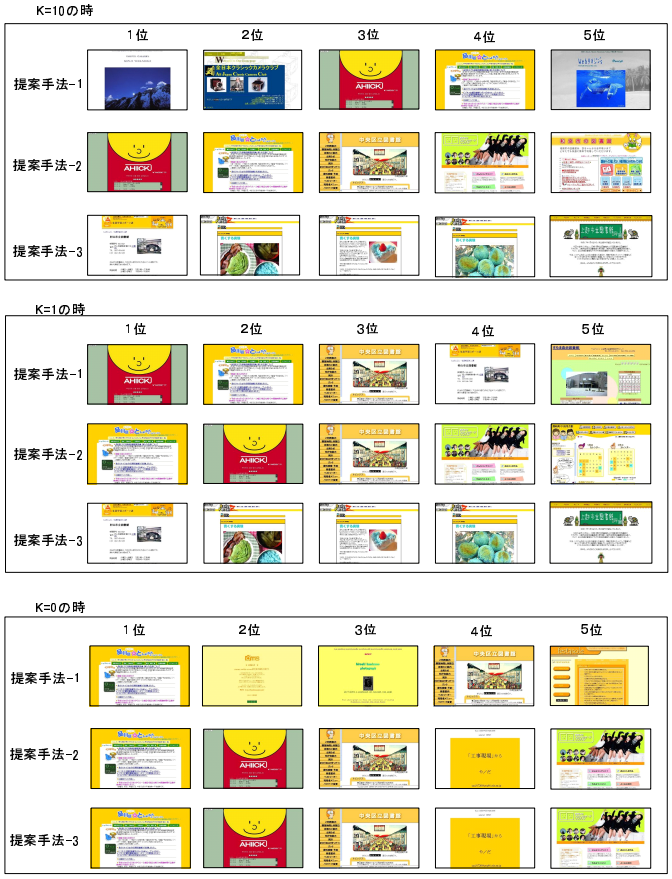

図 9　k の値によるシステムの出力結果の変化（上が黄の Web ページを検索）

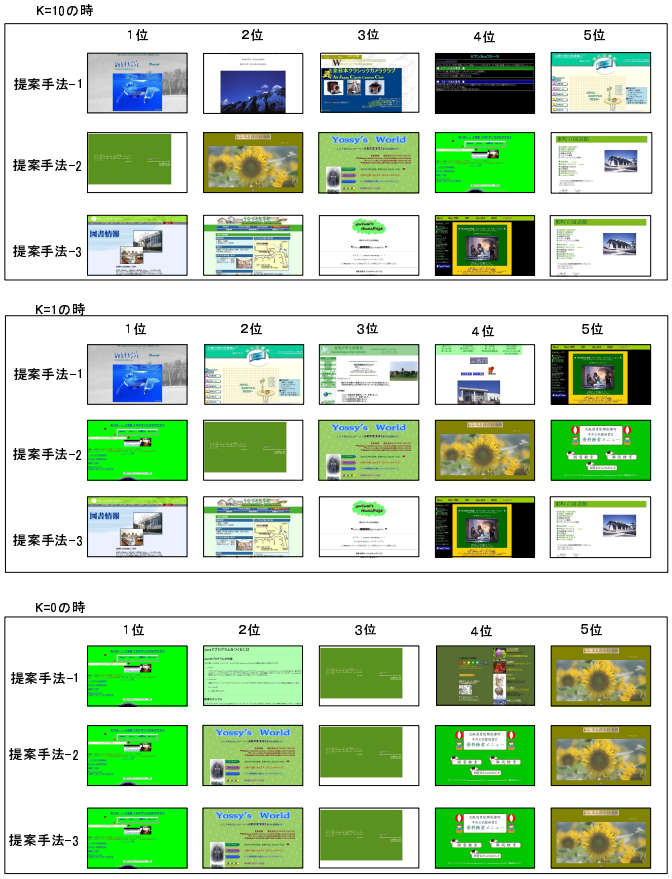

図 10　k の値によるシステムの出力結果の変化（上が緑の Web ページを検索）

表 4　黄色をキーとした Web ページ検索時の適合率

(k = 10)	提案手法-1	提案手法-2	提案手法-3
上位 10 件	22.50%	46.67%	55.00%
上位 20 件	18.75%	45.00%	45.00%
上位 30 件	23.33%	41.11%	44.17%
(k = 1)	提案手法-1	提案手法-2	提案手法-3
上位 10 件	27.50%	36.67%	37.50%
上位 20 件	21.25%	38.33%	43.75%
上位 30 件	21.67%	35.56%	41.67%
(k = 0)	提案手法-1	提案手法-2	提案手法-3
上位 10 件	12.50%	20.00%	22.50%
上位 20 件	17.50%	31.25%	31.25%
上位 30 件	16.67%	34.17%	33.33%

表 5　緑色をキーとした Web ページ検索時の適合率

(k = 10)	提案手法-1	提案手法-2	提案手法-3
上位 10 件	10.00%	44.00%	50.00%
上位 20 件	12.50%	41.00%	46.25%
上位 30 件	10.83%	50.00%	53.33%
(k = 1)	提案手法-1	提案手法-2	提案手法-3
上位 10 件	17.50%	32.50%	27.50%
上位 20 件	18.75%	38.75%	42.50%
上位 30 件	17.50%	50.83%	48.33%
(k = 0)	提案手法-1	提案手法-2	提案手法-3
上位 10 件	20.00%	20.00%	22.50%
上位 20 件	16.25%	32.50%	33.75%
上位 30 件	14.17%	40.83%	36.67%

引色として “白” “黄緑” を多く含む Web ページが上位にランキングされ，検索者は不適合としている．このことから，本実験においては人が理解しやすい色に近づけることを目的として色鉛筆の色を索引色に利用したが，検索キーとして与える色に適した索引色を設定することで適合率が改善すると考えられる．

4. 関 連 研 究

文献 [3], [5], [6] は見た目を用いた Web ページ検索手法である．文献 [3] は Web ページのレイアウトに注目し，検索者がスケッチしたレイアウトを元に類似する見た目の Web ページを検索する手法を提案している．文献 [3] では Web ページレイアウトを参考にすることが目的で設計されているため，スケッチにはテーブル，テキストといった HTML タグの情報を入力する必要がある．文献 [5] は Web ページを画像化してエッジ検出を行うことによりブロックを抽出し，ブロックの配置パターンによって Web ページを比較している．文献 [6] はキーワードによる検索結果から検索者が選択した Web ページと視覚的に類似した Web ページを提示する VIPS アルゴリズムを提案している．文献 [5], [6] では Web ページ同士の類似性を評価しており，検索者が検索キーを生成して各 Web ページと比較する点で提案手法とは異なる．

文献 [1], [2] は類似画像検索手法である．文献 [2] は，個人によ

り異なる視覚感性を階層構造を用いることでモデル化し，類似画像検索へ応用する手法を提案している．文献 [1] は，L*a*b* 表色系を用いて計算した色差を元に，画像に対して感性語を自動抽出する手法を提案している．提案手法は，Web ページの特徴を考慮し，配置位置に対して重み付けを行う点が異なる．

5. おわりに

本稿では，Web ページに含まれる色とその配置位置に基づいた視覚的印象による Web ページ検索手法を提案した．ここで，色とその配置位置に基づいた視覚的印象とは，Web ページがブラウザに表示された直後のイメージのうち，検索者が記憶する部分を指す．本手法の主要な特徴は，Web ページ画像を視覚的印象として，Web ページ検索のための索引自動構成機構を実現する点にある．これにより，検索者は色とその配置位置を記憶している場合でも，検索キーを言葉で与えることが困難な状況において Web ページ検索が可能となった．さらに，Web ページ画像から抽出される特徴を検索者が指定する機構を実現した．これにより，同一 Web ページに対する異なる検索者の記憶の相違による検索精度の低下の抑制が可能となった．本手法のプロトタイプを実装し，実験により本手法の有効性を検証した．

今後の課題は，複数の色を同時に指定すること，言語データを検索キーに用いた検索手法との組み合わせにより精度向上を図ることがあげられる．また，Web ページに出現する色の頻度に合わせた索引色の生成手法も考えていく予定である．

文　献

[1] 小屋夕介，中西崇文，北川高嗣：L*a*b*表色系を利用した静止画像からのメタデータ自動抽出方式，電子情報通信学会 DEWS2003(2003)

[2] 多田昌裕，加藤俊一：階層的分類を用いた視覚感性のモデル化と類似画像検索への応用，情報処理学会論文誌，Vol.44, No.SIG 8, pp.37–45(2003)

[3] 橋本泰成，五十嵐健夫：レイアウトによる WEB ページ検索，インタラクション 2004，情報処理学会ヒューマンインタフェース研究会 (2004)

[4] 福村真哉，春本要，下條真司，西尾章治郎：ユーザの視聴傾向に基づく Web コンテンツ個人化提示システム，情報処理学会論文誌，Vol.45, No.SIG 14, pp.12–22(2004)

[5] 三橋憲晃，山口亨，高間康史：視覚的類似性に基づく Web ページ検索手法の提案，第 17 人工知能学会全国大会 JSAI2003(2003)

[6] Shipeng Yu, Deng Cai, Ji-Rong Wen, Wei-Ying Ma：Improving Pseudo-Relevance Feedback in Web Information Retrieval Using Web Page Segmentation, Proceeding of the 12th International World Wide Web Conference(2003)

[7] 越智宏，黒田英夫：JPEG & MPEG 図解でわかる画像圧縮技術，日本実業出版社，pp.68–71(1999)

[8] 日本規格協会：JIS ハンドブック（色彩 2004），日本規格協会 (2004)

[9] 名阪カラーワーク研究会 http://www005.upp.so-net.ne.jp/fumoto/linkp25.htm