

誘目性を活用した画像加工による注視誘導

萩原 愛子[†] 杉本 晃宏^{††} 川本 一彦[†]

[†] 千葉大学大学院融合科学研究科 〒263-8522 千葉県千葉市稲毛区弥生町 1-33

^{††} 国立情報学研究所 〒101-8430 東京都千代田区一ツ橋 2-1-2

E-mail: [†]{hagi,kawa}@chiba-u.jp, ^{††}sugimoto@nii.ac.jp

あらまし 誘目性を測る指標として視覚的顕著性が注目されている。そして、画像中の指定領域の視覚的顕著性を高めることで、人の注視を自然に誘導する手法が提案されている。そこでは、指定領域内の視覚的顕著性を高め、領域外の視覚的顕著性を下げするために、入力画像の各ピクセルのRGB値を独立に変更する操作によって画像を加工している。そのため、注視自体は誘導できるものの、加工画像と入力画像との間に違和感が生じるという問題がある。そこで、本稿では、画像加工によって生じる違和感を軽減するために、画像の加工に画像の輝度値を補正する操作を取り入れることを考える。具体的には、入力画像と加工画像との間で色ヒストグラムになるべく差が出ないように輝度値の補正を画像全体で一括して行う。これによって、指定領域の顕著性を高めつつ違和感のない画像加工を実現する。被験者実験によって、提案手法の有効性を確認したので、それも併せて報告する。

キーワード 注視, 視覚的顕著性マップ, 画像加工, 画像補正, 誘目性, 視線インタフェース。

Saliency-Based Image Modulation for Gaze Navigation

Aiko HAGIWARA[†], Akihiro SUGIMOTO^{††}, and Kazuhiko KAWAMOTO[†]

[†] Graduate School of Advanced Integration Science, Chiba University Chiba, Japan

^{††} National Institute of Informatics Tokyo, Japan

E-mail: [†]{hagi,kawa}@chiba-u.jp, ^{††}sugimoto@nii.ac.jp

Abstract Saliency map is promising for index to measure the gaze-attention value. The method for editing an image, when given a region in the image, to synthesize the image in which the region is most salient is proposed. In that method, each RGB channels are modulated independently to control the saliency. The intensity and color are modulated until the saliency inside the region becomes the highest for the entire image. Therefore, change of appearance between input and modulated image provides a feeling of strangeness, even though gaze are navigated to the given region. In response to this, we propose to include the image correction step in the image modulating process, so that to alleviate a feeling of strangeness. Specifically, correct the brightness value for the entire image in one lump to close the gap between the histograms of input and modulated images. These will allow to realize image modulating method, which raise the saliency of a given region without providing a feeling of strangeness. We also report on experimental results confirm that our image modulation method naturally draws the human visual attention toward our specified region.

Key words visual attention, saliency map, gaze-based interface.

1. はじめに

我々は、人間の活動を支援する情報システムに囲まれている。事実、日常では携帯電話をはじめとして、街頭の音声案内や電子掲示板などといった情報システムを頻繁に利用している。このように、情報システムは、もはや、我々の生活に欠かせないものとなっている。誰もが日常的に利用するものであるからこ

そ、これらは人間との自然なインタラクションを通して利用可能であることが求められている。

情報システムを対象としたインタフェースの手段は主に手や音声や顔である。なかでも、視線を利用したインタフェースが特に注目されている。視線は他の手段に比べ、ユーザの意図や注意をよりよく反映すると考えられるからである。自然に注視を誘導する視線インタフェースがあれば、ユーザにかかる負担

も少なくなり、情報システムをより容易に利用することが可能となる。また、将来的にはウェアラブルディスプレイと組み合わせた道案内や情報提供などにも応用できると考えられる。進行方向や有益な情報がある場所を文字や記号で明示するのではなく、そこに自然と視線が向くようにすることによってユーザビリティを高めることができると考えられるからである。

視野内に刺激を呈示し、注視を誘導する方法がある [6]。周辺視野内に刺激が呈示されると、現在の注視を中断し、視線を移動させて刺激部分を新たに注視することになる。LED の点灯を利用した注視誘導システム [8] もこれを利用している。しかしこういった方法は、警告としては有意義だが、現在注視している対象を強制的に変えるという面で人の行動を阻害する要因を含んだ注視誘導になっている。一方、周辺視野へ視覚刺激を呈示し、サッカード中にその刺激を消失させることで利用者に気付かせずに視線を誘導する SGD と呼ばれる手法がある [1]。このような手法によって、周辺視野への視覚刺激呈示により、探索タスクの時間が短くなることが確認されている [5]。これは視線誘導という点では確かに有効だが、常に利用者の視線位置を測定する必要があり、なおかつ同時に二人以上の視線を誘導することができない。その他にも、アイコンタクトや指差しによって視線を誘導することも可能である [9]。しかしこれは利用者が指差しといった指標を最初から注視していることが前提となっている。この指標部分への注視の誘導が確立できていない限り、利用することができない。また、幾何学模様のエッジを強調することで、強調した部分へ視線を誘導することも可能だが [4]、これは利用できる画像の種類が限られている。

一方、無意識かつ自然に注視を誘導するため、視覚的顕著性 [3] を利用する方法が提案されている [2]。視覚的顕著性とは人間の注視の引きつけやすさを意味し、画像から単純な視覚特徴を抽出・統合することによって、人間が注意を向けやすい画像中の領域を視覚的顕著性マップとして求めることができる。そこでは、視覚的顕著性が高いほど、人間の注視を引きつけやすいとされている。ユーザの興味や意図といった要因を考慮しなければ、人間は視野内で最も視覚的顕著性の高い領域を注視すると考えられている。画像中で指定した領域の視覚的顕著性が最も高くなるように入力画像を加工している。この手法では入力画像に対して、視覚的顕著性が指定領域内で高くなり、領域外で低くなるように明るさと色を繰り返す変化させる。そして、指定領域内の視覚的顕著性が画像内で最も高くなった時点での画像を出力する。

ところが [2] で提案されている手法では、視覚的顕著性を操作する為に RGB 値を独立に変更させているため、加工後画像の見えが入力画像の見えと大きく異なるという不具合が生じる可能性がある。加工によって、指定領域への注視発生確率が高くなったことは確認されているが、加工前後の画像値の差分の大きさや、それに伴う見えの違和感については言及されていない。

そこで本稿では、画像加工によって生じる違和感を軽減させた視覚的顕著性に基づく画像加工方法を提案する。具体的には、指定領域へ注視を誘導するために独立に変更された RGB 値に

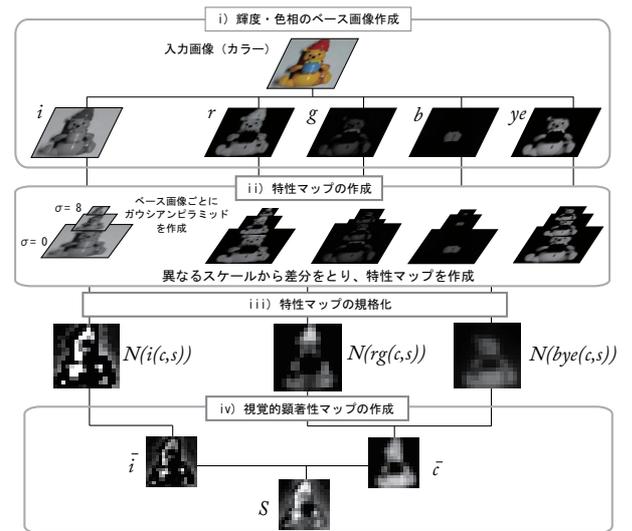


図 1 視覚的顕著性マップの計算過程の概要

対して、周辺と馴染むよう変化量に平滑化を施す。さらに、入力画像と加工画像と間の色ヒストグラムになるべく差が出ないように輝度値の補正を画像全体で一括して行う。このような操作によって、画像加工による違和感を軽減させながら、指定領域の顕著性を高めることを同時に実現させる。

2. 視覚的顕著性に基づく画像加工

画像中で指定した領域の視覚的顕著性が最も高くなるように入力画像を加工する手法が提案されている [2]。また、同手法を用いた画像加工によって、指定した領域への注視発生確率が上昇することも確認されている。この手法ではカラーの入力画像に対して、視覚的顕著性が指定領域内で高くなり、領域外で低くなるように明るさと色を更新することを繰り返している。そして、指定領域内の視覚的顕著性が画像内で最も高くなった時点での画像を出力としている。 $\gamma_{xy}(\gamma = R, G, B)$ を各 (x, y) 座標における更新前の画素値、 γ'_{xy} を更新後の画素値として、一度の更新方法は次のように定義されている

$$\gamma'_{xy} = \gamma_{xy} + p_{xy} Q_{\gamma_{xy}}(\gamma = R, G, B). \quad (1)$$

ここで、 p_{xy} は一度の更新で行う変化の度合いを決定する強度係数である。これは次のように求める。まず、更新前の画像の視覚的顕著性を全ピクセルに対応させて与える。こうして、視覚的顕著性の高い点は強度係数の絶対値が大きく、低い点は小さくなる。ただし、視覚的顕著性を下げるべきである領域外の $(x, y) \notin D$ であれば、そこでは与えられた視覚的顕著性を負の値とする。最後に全体にガウシアンフィルタをかけ平滑化し、各ピクセルがもつ値をそのピクセルの強度係数 p_{xy} と定義する。また、 $Q_{Rxy}, Q_{Gxy}, Q_{Bxy}$ は画素ごとの変化量である。これは視覚的顕著性マップの計算過程において視覚的顕著性が輝度や色相のどの特徴量の影響を受けているかを判別し、 R, G, B それぞれに適応して値を定める。視覚的顕著性マップの計算過程の概要を図 1 に示す。

たとえば、周囲と比べ赤みが強い点は、赤の色相が視覚的顕著性に大きな影響を与えているため、 Q_{Rxy} は大きな値をも

表 1 エッジからの距離とガウシアンカーネルの半径

距離	半径
エッジ上から 3pixel まで	0pixel
4-5pixel	3pixel
6-8pixel	5pixel
8pixel 以上	7pixel

つ. なお, 「周囲」とは色と明るさごとの画像全体の平均値と定義されているが, 本手法では「周囲」を次のように定義する. まず, 視覚的顕著性マップの計算過程において, 各座標ごとにどのガウシアンピラミッドの画像が最も大きな値を持っていたのか比較する. そのときの画像の粗さを示すスケールの s ($s = 5, 6, 7, 8$) 値に応じて, 各座標を中心とした $2^s * 2^s$ 内の色と明るさごとの平均値をその点の周囲の値と定義する.

3. 画像の見え補正

制御後の画像に対して, 画像の見えを向上させる為の見えの補正方法を提案する.

3.1 輝度値変化量の補正

画素ごとの変化量と強度係数の積によって定義された変化量は各画素で独立して定義されており, 近隣の画素値の変化量が考慮されていない. それでは, 隣接した画素同士であっても大きく異なる変化量が設定され, 更新後の画像の見えに違和感を与える可能性がある. そこで, 周辺の変化量を考慮する為に, $p_{xy}Q_{\gamma_{xy}}$ に対して平滑化処理を施す. ただし, たとえ近い範囲にある画素であってもそれがエッジの境界を挟んでいた場合, 平滑化を行うのは望ましくない.

まず, 更新前画像からエッジを検出する. そしてそのエッジからの距離に応じて, ガウシアンフィルタのカーネルの大きさを表 1 に示すように定義する.

なお, エッジは 4 段階の細かさを用意し, 各画素の視覚的顕著性に与えているスケールの大きさごとに, 利用するエッジを選択する. こうした平滑化処理を $Gau(\cdot)$ と記し, $p_{xy}Q_{\gamma_{xy}}$ に対して平滑化したものを $Gau(p_{xy}Q_{\gamma_{xy}})$ と記す.

3.2 輝度値の補正

視覚的顕著性を制御する為に画像の見えが入力画像と大きく異なることがある. 加工による画像の見えの違和感を軽減するには, 制御後の画像を入力画像に近づけるように補正することが必要である. それは輝度値にバイアスを設定したり, ダイナミックレンジを補正することで, 制御後画像の見えを入力画像に近づけられると考えられる. その効果を図 2 に示す. 加工前の画素値を R , 加工後の画素値を R' としたとき, 加工前後の画素値の変化量が少なければ図 2 (a) に示すように, 各画素から得られる輝度値の分布が直線 $R=R'$ 付近に集中することになる. 加工によって, 直線 $R=R'$ 輝度値の分布が大きく離れてしまっている図 2 (b) のような状況になった場合バイアスの設定とダイナミックレンジ補正を施すことで, 図 2(c) のように輝度値の分布を直線 $R=R'$ に近づけることができれば, 加工前後の画素値の変化量が小さくなる. このとき, 個々の画素単位で補正してしまうと, 制御した視覚的顕著性までも入力画像の

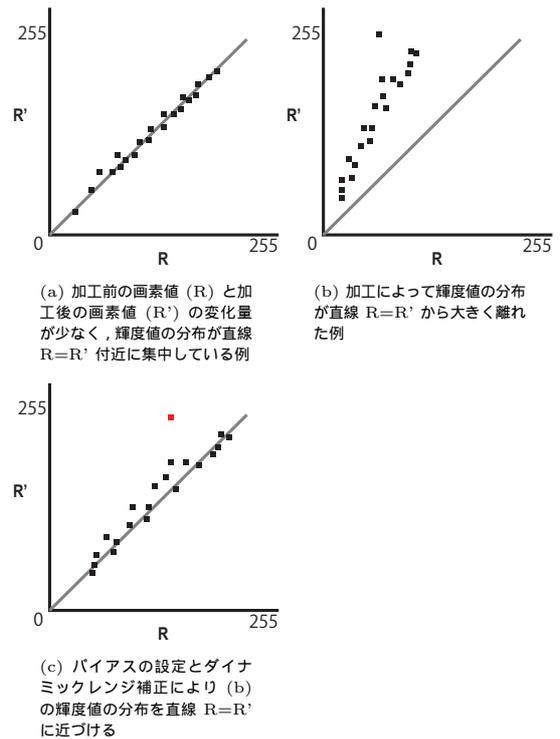


図 2 補正による効果

ものに戻ってしまう可能性があるため, この補正は画像全体に対して一括した処理を施すべきである.

画素値にかかるバイアスや比率を求めることは, 入力画像の画素値を γ_{xy} , 更新後の画素値を γ'_{xy} としたときに, その差の総和を最小とさせる乗算と加算方法を求める最小二乗誤差による直線当てはめ問題と考えることもできる. 最小二乗誤差による直線当てはめ問題は, 図 2 (c) において赤色で示したような外れ値の影響を大きく受けやすいと言われているが, 各画素に対して輝度値をどれだけ変化をさせるかがわかっているため, どの画素が外れ値になり得るかは本問題では既知である. そこで, 外れ値になり得る画素には小さい重みを与え, 外れ値になり得ない画素には大きい重みを与えて, 重みつき最小二乗誤差を評価することで, 外れ値の影響を受けない当てはめを実現することが可能となる.

式 (1) より, 一度の更新による画素値の変化量の大きさは強度係数 p_{xy} によって定まることがわかる. よって, $|p_{xy}|$ が大きいほどその座標は外れ値をなりやすいといえる. この性質を利用することによって, 外れ値の影響を軽減させることができる. 各座標の重み w_{xy} は, 各座標の強度係数 $|p_{xy}|$ と強度係数の最大値 p_{MAX} を利用して次のように定義する,

$$w_{xy} = \frac{p_{MAX} - |p_{xy}|}{p_{MAX}}$$

そして, この重みを用いて,

$$\sum_{xy} w_{xy} \{ \gamma_{xy} - (\alpha_{\gamma} \gamma_{xy} + \beta_{\gamma}) \}^2$$

を最小とする $\alpha_{\zeta}, \beta_{\zeta}$ ($\zeta = R, G, B$) を計算する. 次に, 更新後の画素値 γ'_{xy} に対して $\alpha_{\zeta}, \beta_{\zeta}$ を用いた補正を次のように施し, 補正後の画素値 $\widehat{\gamma'_{xy}}$ を得る.

$$\widetilde{\gamma'_{xy}} = \alpha_{\zeta} \gamma'_{xy} + \beta_{\zeta}.$$

具体的な画像更新の手法を Algorithm 1 に記す.

Algorithm 1 画像更新方法

- 1: 顕著性の制御 $\gamma'_{xy} = \gamma_{xy} + Gau(p_{xy} Q_{\gamma_{xy}})(\gamma = R, G, B)$
 2: 重み付き最小二乗正規化 $\widetilde{\gamma'_{xy}} = \alpha_{\zeta} \gamma'_{xy} + \beta_{\zeta}$
-

4. 実験

4.1 目的

提案手法による補正を含む加工と、補正の無い加工に関して、加工画像の見えの違和感を評価する。また、補正の導入が注視誘導へ与える影響を、被験者を用いた実験によって検証する。

4.2 手順

4.2.1 画像加工

- (i) 512 × 256 のカラー画像 In 40 種を入力とし、視覚的顕著性マップを計算する。
- (ii) 視覚的顕著性を上げる領域 D を指定する。なお、入力画像 40 種のうち 10 種には、2 種類の異なる領域を指定する。
- (iii) 一つの入力に対して、二通りの更新を並列に行う。Algorithm 1 を適用した補正有りの更新と、補正を施さない更新を行う。その際、更新のたびに視覚的顕著性マップを求める。視覚的顕著性の最も高い座標が領域内の座標となるまで更新を繰り返す。
- (iv) 指定領域の顕著性が最も高くなった画像を出力する。補正有りの画像は OutA、補正無し画像は OutB とする。

4.2.2 画像検証

補正の有る OutA と補正の無い OutB に関して、次の値を測定することで比較検証を行う。

- (i) 加工後画素値と入力画素値の差分の絶対値。
- (ii) 入力画像と加工後画像のヒストグラムにおける Bhattacharya 距離。
- (iii) 入力画素値に対する 加工後画素値の常用対数の絶対値。

4.2.3 被験者実験

TobiiTX300 を用いて 24 人の被験者の注視点を観測する。ディスプレイと被験者の眼球の距離はおよそ 60cm とし、0.1 秒間に 0.9 度以上の眼球回転がなければ、その点を注視したとみなす。

40 種の入力画像 In と 50 種ずつの加工後の画像 OutA, OutB を 3 秒ずつランダムに提示する。画像と画像を切り替える際には画面中央に十字を 1 秒間提示する。被験者には十字が表示されたら十字の位置へ注視を向けるように、画像が表示されている間は自由に画像を眺めるように指示する。画像を提示していた間に、一度でも指定領域内への注視が発生したかどうかを測定する。

4.3 結果

入力画像とその顕著性マップ、指定領域、加工後の画像とそ

(注): 画像: NHKクリエイティブ・ライブラリー



図 3 In1^(注)と D1 による加工結果



図 4 In2^(注)と D2 による加工結果

の顕著性マップを図 3, 4, 5, 6 に示す。図 3, 4, 5, 6 の (d) と (e) は、画像全体のうち最も高い顕著性を示す点が領域内にあるという点は等しいが、画像としての見えが全く同じになると



図 5 In3^(注)と D3 による加工結果



図 6 In4^(注)と D4 による加工結果

いうわけではない。図 3, 4, 5, 6 の (d), (e) と (a) を比較すると、補正をしなかった場合、加工によって画像全体の明度が入力画像から異なりやすいということがわかる。

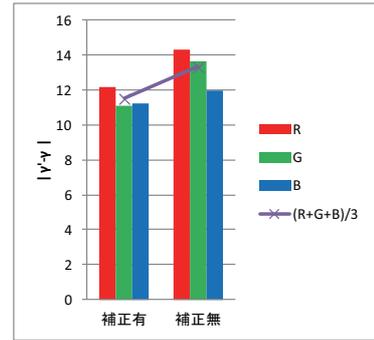


図 7 加工後画素値 (γ') と入力画素値 (γ) の差の絶対値 ($\gamma = R, G, B$)

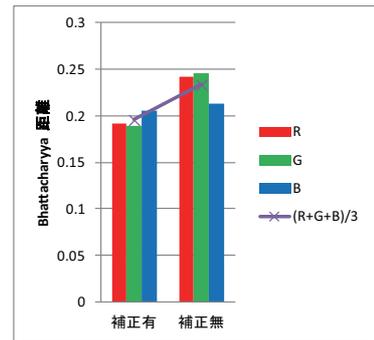


図 8 入力画像と加工後画像の Bhattacharyya 距離

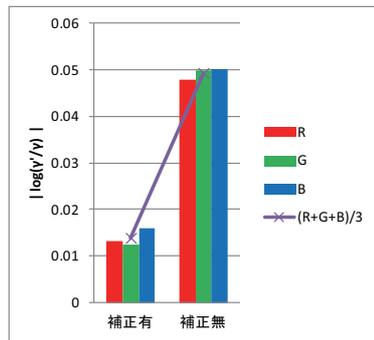


図 9 入力画素値 (γ) に対する 加工後画素値 (γ') の常用対数の絶対値

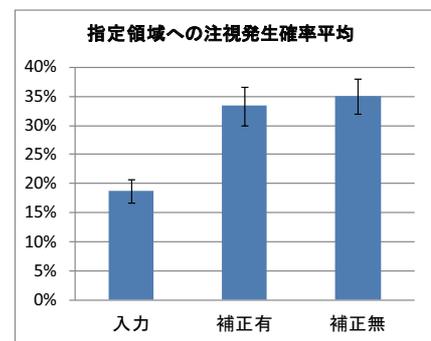


図 10 画像種類ごとの指定領域への注視発生確率

入力画像と加工後画像の画素値の差の絶対値の平均値を図 7 に示す。この値が小さい程、加工前後の画素値の変化量が小さい。補正有りとしを比較すると、見えの補正を行うことで、加工後の画素値を入力画像の画素値に近づける効果があることがわかる。

入力画像と加工後画像のヒストグラムの Bhattacharya 距離を図 8 に示す。Bhattacharya 距離は二つのヒストグラム間の類似度を表し、ヒストグラムが完全に一致した場合は 0 となる。補正の有無で比較すると、補正有りのほうが数値が低く、入力画像と類似したヒストグラムとなっていることがわかる。

入力画素値に対する加工後画素値の常用対数の絶対値を図 9 に示す。ウェーバー・フェフィナーの法則 [7] より、刺激の強度に対してそれを知覚する感覚量は強度の常用対数に比例すると言われている。入力画像の画素値に対する加工後画像の画素値の常用対数が小さいほど、加工によって人が知覚する感覚量も小さいと考えられる。補正の導入によってこの感覚量が小さくなり、画素値の差分だけでなく、人の知覚レベルにおいても補正によって違和感が軽減されたことがわかる。

入力画像と加工後画像それぞれに対して、指定領域への注視が一度でも発生した確率を図 10 に示す。入力画像の注視発生確率 18.8% に対して調整有りは 33.3%、調整無しは 35.1% となっている。入力画像と比較すると、補正の有無によらず指定領域への注視の発生確率は増加したことがわかる。しかし、補正の有無に関して比較すると、見えの補正を導入したことによって、注視の発生確率は約 2 ポイント減少するという影響を受けた。

4.4 考察

図 7, 8, 9, 10 より、補正を施すことによって、加工後画像の見えは入力画像に近づくものの、注視の発生確率は僅かではあるが減少している。視覚的顕著性マップの理論においては画像の中で高い顕著性を示す点に注視が誘導されやすいとされている。画像の中で最も高い顕著性を示すという点が等しければ、補正の有無に寄らず注視の発生確率は等しくなりそうだが、異なる結果となった。画像内で最も高い顕著性を示すといっても、加工方法が異なれば視覚的顕著性マップは別のものとなり、領域外において高い顕著性を示す他の点との位置関係といった条件も変わってくる。約 2 ポイントの減少という数値が有意であるかという点に関して議論の余地はあるが、注視発生確率の差はこうした条件の違いによって生じたのではないかと推測できる。

画像の補正においてはヒストグラムの性質を直接利用してはいないが、画像全体の画素値にバイアスを設定しダイナミックレンジを揃えるという処理によって、結果的にヒストグラムの類似度という面でも入力画像に近づいたことになる。

図 3, 4, 5, 6 より、指定領域が最も高い顕著性を示すという点は同じであっても、加工後の画像の見えは補正の有無によって差があることがわかる。特に、図 3, 4 に関しては画像中央部分の空の明るさに大きな差があることは気付きやすい。

図 7, 8, 9 の結果を踏まえ、入力画像と見比べると、補正をすることで画像全体の見えを入力画像に近づけられたことは確かである。しかし、これは一例の補正方法に過ぎない。視覚的顕著性マップに対応する画像が一意に定まらないということからもわかるように、より見えの違和感を軽減させる補正方法の検討も望まれる。図 10 より、制御した視覚的顕著性を維持できれば、画像の見えを補正しても注視を誘導できることはわかっ

た。これより今後は、加工後画像の見えをより入力画像に近づけることで、見ている人に察知させることのない注視誘導を実現させられると考えられる。

5. おわりに

人の注視を遮ることなく、自然と視線を誘導する視線インタフェースの実現を目指して、本研究では、画像加工によって生じる違和感を軽減した視覚的顕著性を制御する画像加工を提案した。視覚的顕著性を制御する為に輝度と色相を画素ごとに変化させた後、エッジの隔たりのない領域内に関しては、その変化量が周辺画素と馴染むように平滑化処理を行った。また、入力画像と加工後画像の色ヒストグラムが近づくように画像全体の画素値に一括したバイアスを設定し、入力画像とダイナミックレンジが揃うように補正を行った。こうして視覚的顕著性制御のための変化と見えの補正を繰り返し、人の注視を誘導させたい領域が最も高い顕著性を示すように画像を加工した。補正の有無による見えの違和感の差を、加工後画像を比較することで検証した。画素値の差分やヒストグラムの類似度という点から、補正によって加工前後の画像の見えの変化は小さくなったことを検証し、提案手法の有効性を確認した。また、被験者の注視点を測定した結果、入力画像と比較すると注視の発生確率は補正の有無によらず増加することが確認できた。

謝辞

本研究の一部は、JST CREST「共生社会に向けた人間調和型情報技術の構築」領域採択課題「日常生活空間における人の注視の推定と誘導による情報支援基盤の実現」により実施した。

文献

- [1] Reynold Bailey, Ann McNamara, Nisha Sudarsanam, and Cindy Grimm. Subtle gaze direction. *ACM Trans. Graph.*, 28(100):1-14, 2009.
- [2] A. Hagiwara, A. Sugimoto, and K. Kawamoto. Saliency-based image editing for guiding visual attention. In *Proceedings of the 1st international workshop on Pervasive eye tracking & mobile eye-based interaction*, pages 43-48. ACM, 2011.
- [3] L. Itti and C Koch. Computational modeling of visual attention. *Nature Reviews Neuroscience*, 2:194-230, 2001.
- [4] Youngmin Kim and Amitabh Varshney. Persuading visual attention through geometry. *IEEE Transactions on Visualization and Computer Graphics*, 14:772-782, 2008.
- [5] Ann McNamara, Reynold Bailey, and Cindy Grimm. Improving search task performance using subtle gaze direction. In *Proc. of the 5th symposium on Applied perception in graphics and visualization (APGV'08)*, pages 51-56, 2008.
- [6] 緒方 康匡 and 内川 恵二. 第一サッカードを誘導するための視覚刺激条件. *ITE technical Report*, 33(17):57-60, 2009.
- [7] 日本視覚学会編. 視覚情報処理ハンドブック. 朝倉書店, 2000.
- [8] 猪目 博也, 饗庭 絵里子, 下斗米 貴之, 藤澤 隆史, and 長田 典子. Led 警光灯の視認性向上: 目立ちやすさの感性指標に基づく点滅パターン評価法. *映像情報メディア学会技術報告*, 34(10):89-92, 2010.
- [9] 鈴木 祐也, 葛岡 英明, 山下 淳, 山崎 敬一, 山崎 晶子, and 久野 義徳. 無言のロボットによる注意誘導の研究. *ヒューマンインタフェースシンポジウム 2007*, pages 615-618, 2007.