身体への印象単語を判別する際に計測された視線位置分布の確率表現 *

木 下 顕 **，井 上 路 子 **，西 山 正 志 **，岩 井 儀 雄 **

Probability Representation of Gaze Distribution Measured When Judging Impression Words of Body Parts

Ken KINOSHITA, Michiko INOUE, Masashi NISHIYAMA and Yoshio IWAI

We investigate how to represent the probability of gaze distributions for indicating that the observers frequently view body parts when they judge impression words for the body parts of the subjects in the person images. In the field of cognitive science, analytical studies have been reported on how observers view person images and judge the impressions of the subjects. However, there was no discussion of how to represent the probability of gaze distributions when judging impressions for the subjects. Our method gives the observers the task of judging impression words related to a formal scene, and measure the gaze locations from the observers. We represent a conditional gaze probability of body parts using the measured gaze locations. We evaluated how the gaze probabilities change between impression words and body parts included in the tasks. We confirmed that there was a tendency of causing differences among impression words because the divergences between the conditional gaze probabilities of the body parts were large.

Key words: Impression words, body parts, observers, gaze distribution, probability

1. はじめに

結婚式やパーティーなど，大勢の人間が社会的な通念を共有できるフォーマルなシーンにおいて，人物の印象は重要となる．実際に我々が知人の結婚式に出席する際，そのシーンに相応しい印象を出席者に与えることは大事である．本論文では，印象を推定するシステムを考えるために，そのシステムの目となるカメラで撮影された人物画像を対象とする．フォーマルなシーンで撮影された画像中の被写体から受ける印象を以下で取り扱う．なお，観察者と画像中の被写体は，互いに初対面とし議論を進める．

フォーマルなシーン中に存在する人物の印象を表す単語として，美しい，可愛い，清潔感がある，上品である，素敵であるなどが挙げられる．これらの印象単語について，人物画像を見ている観察者に尋ねると，多くの観察者は共通の回答を思い浮かべると考えられる．例えば，Fig. **1** (a) の人物画像について，美しいか否かを観察者へ尋ねる．この場合，多数の観察者が美しいと回答し，一部の観察者が美しくないと回答すると想定される．次に，美しいとは別の印象単語を，異なる人物画像について尋ねた場合を考える．清潔感があるか否かについて尋ねると，先ほどと同様の傾向が得られると想定される．その例を Fig. **1** (b) に示す．

ここで，画像中の被写体に対する印象単語を，システムが自動で推定する技術について考える．人物画像から印象単語を推定するために，一般的には，機械学習や深層学習を適用することが多い．人物画像と印象単語のペアを収集することで，画像と単語の相関関係を統計的に学習していく．これらの機械学習や深層学習の技術は，様々なアプリケーションで活用されている．例えば，人物の性別や服装などの属性認識 [1, 2]，写真の審美品質識別 [3, 4] において活用事例が報告されている．人物画像

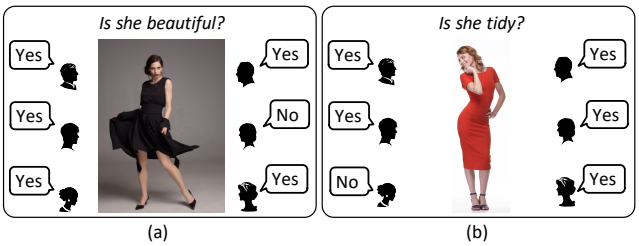

Fig. **1** Examples of questions using impression words for person images.

と印象単語のペアからなる訓練サンプルを大量に収集することができれば，機械学習や深層学習の技術を用いることで，印象推定のアプリケーションにおいても高い精度を得られると期待される．ただし，訓練サンプルを大量に収集することは多大な手間がかかる課題があった．

観察者から計測された視線位置分布を学習アルゴリズムへ組み込むことで，訓練サンプルのみに頼らずに機械学習や深層学習の精度を高める手法 [5, 6, 7, 8, 9, 10, 11] が近年登場しつつある．これらの既存手法は，大量の訓練サンプルが収集できない場合でも，精度向上への寄与が期待される．ただし，これらの既存手法では，印象単語に対する観察者の視線位置分布を，機械学習や深層学習において，どのように取り扱うかについては議論されていなかった．

本論文で取り扱う視線位置分布を明確にするために議論を進める．認知科学の分野において，観察者が人物画像をどのように見て，その被写体の印象を持ったかについての分析研究 [12, 13, 14, 15, 16] が報告されている．これらの分析研究では，人物画像に対して印象を評価するタスクを観察者に与え，その観察者の視線を計測している．観察者は，顔を主として様々な身体部位へ，視線を空間的に配ることが明らかにされている．また，タスクに含まれる印象単語によって，視線の配り方が変化することが明らかにされている．ただし，これらの分析研究では，印象を判別するタスクを与えた時の視線の配り方を，機械学習や深層学習に組み込む考え方については議論されていな

* 原稿受付 令和 2 年 5 月 8 日
* 掲載決定 令和 2 年 7 月 30 日
** 鳥取大学大学院工学研究科 (鳥取市湖山町南 4 丁目 101)

かった．確率統計に基づく機械学習や深層学習のアルゴリズムへ，観察者の視線位置分布を円滑に組み込むためには，身体部位への視線の配り方そのものが確率的に表現されていることが望ましい．

そこで本論文では，画像中の被写体の身体部位に対する印象単語を観察者が判別する際に，どの身体部位をどれだけ見たかを表す視線位置分布を，確率で表現するための手法について述べる．また，実際に視線位置を計測し，印象単語の間で確率分布に違いが表れるかどうかを調査する．提案手法では，観察者の視線位置を計測するために，フォーマルなシーンで用いられる印象単語を判別するタスクを観察者に与える．計測された視線位置を，身体部位を条件として与えた時の確率分布として表現する．実験において，タスクに含まれる印象単語の間で，視線位置分布がどのように変化するかを比較した．フォーマルなシーンにおける印象単語の間で，身体部位に対する視線の空間的な配り方に差異がある傾向が見られることが，条件付き確率の分布間の距離から分かった．なお本論文では，機械学習や深層学習のアルゴリズムへ視線位置分布を組み込むレベルまで到達できてはいないが，そのための重要な要素技術となる視線位置分布の確率表現を開発した点に主な貢献がある．以下，2. で関連研究について述べ，3. で視線計測について述べ，4. で視線位置分布の確率表現について述べ，5. で実験について述べる．最後に，6. でまとめる．

2. 関連研究

2.1 機械学習や深層学習へ視線位置を組み込む既存手法

いくつかの認識タスクに向けて，実際の観察者から計測された視線位置を，機械学習や深層学習へ組み込む手法が開発されている．文献[5]では，動画からの文章生成を目的として，観察者の視線位置を利用し，文章の予測精度を向上させている．文章生成ネットワークのアテンション機構において，動画中のどこに視線が集まるかを教師信号として取り入れている．文献[6]では，画像に対する質疑応答を目的として，観察者の視線位置を利用し，応答精度を向上させている．先の既存手法と同様に，視線でアテンション機構を制御している．文献[7]では，動画要約を目的とし，視線を利用することでユーザの好みへ合わせることを実現している．具体的には，動画を見ている観察者から計測された視線の位置と速度を，深層学習の入力として組み込んでいる．文献[8]では，ユーザ好みの画像選択を目的とし，視線位置を特徴量として認識アルゴリズムに組み込んでいる．文献[9]では，ファッションアイテム画像の属性認識を目的とし，視線を用いることで認識精度を向上させている．視線が集まりやすい領域をマスキングし，その領域内で学習のための特徴量を抽出している．文献[10]では，どのファッションアイテム画像をユーザが好むかを推定することを目的とし，深層学習のプーリング層へ視線位置を組み込んでいる．文献[11]では，人物画像の性別認識を目的とし，視線位置を用いることで精度を向上させている．特徴抽出の前処理として，観察者の視線が集まる位置へ大きな重みを与えている．これらの既存手法から分かるように，観察者から計測された視線位置は，機械学習や深層学習の精度向上に寄与すると言える．ただしこれらの既存手法では，人物画像の印象を推定するタスクについては議論されていなかった．本論文では，人物画像の印象推定において視線を活用することを目的とし，印象単語を判別した際の観察者の視線位置を，確率分布で表現することを狙う．

2.2 印象に対する視線位置の分析研究

認知科学の分野において，画像中の被写体を対象とし，観察者がもった印象と視線位置との関係について分析研究が報告されている．文献[12]では，男性の体型変化の画像を用いて，好印象と感じるかの多段階評価を行い，視線位置を分析している．被写体の印象を決める際に，観察者は顔へ視線を主に配り，次に胸に視線を配ることが述べられている．文献[13]では，様々な女性の全身画像を用いて，第 1 印象の良し悪しの多段階評価を行い，視線位置を分析している．男性の観察者は，胸や腰と比べて顔に視線を配ることが述べられている．文献[14]では，女性の体型変化の画像を用いて，魅力をどの程度感じたかの多段階評価を行い，視線位置を分析している．男性の観察者の視線は，腰や足と比べて顔や胸に集まることが述べられている．文献[15]では，アルコールを摂取した観察者が持つ印象と視線位置の関係を分析している．他の分析研究の結果と同様に，腰や胸と比べて顔に視線が集まることが述べられている．文献[16]では，化粧販売員の画像を用いて，親近感や信頼感の多段階評価を行い，視線位置を分析している．観察者は，商品を持つ手に比べて顔へ視線を配っていることが述べられている．以上の分析研究より，観察者が印象を決める際，顔を主に見るが，それ以外の身体部位も見ると言える．このことから，顔や手足などの身体部位に向けて，視線を何処へどれだけ空間的に配るかが印象に深く関係すると考えられる．本論文では，これらの既存の分析研究の知見を機械学習や深層学習で取り扱えるように，印象に対する視線の配り方を確率分布で表現することを狙う．

3. 観察者の視線位置の計測

3.1 概要

確率表現された視線位置分布を，機械学習や深層学習に組み込む価値があるかどうかを決めるために，実際に印象を判別する推定処理を構築し，その精度で結論付けることが望ましい．ただし，印象に応じて視線位置分布がそもそも変化しないのであれば，推定精度の向上は見込めない可能性が考えられる．

そこで本論文では，事前調査として，確率表現された視線位置分布が，印象によって変化するかどうかを確認する．この調査のために，人物画像に含まれる身体に対して，その印象を判別するタスクを実験協力者に与え，視線位置を計測する．その際に，以下の仮説を置き検証を進める．

仮説：与えたタスクに含まれる印象単語と身体部位との組み合わせが変化すると，視線位置分布に変化が生じる．

印象単語として，大勢の人間が社会的な通念を共有できるフォーマルなシーンに当てはまるものを対象とする．具体的には，美しい，清潔感がある，可愛いを用いる．また，身体部位として手と足を対象とする．これらの身体部位をタスクに含める理由を以下に挙げる．印象に関する視線の分析研究[12, 13, 14, 15, 16]では，実験協力者が人物画像の印象を感じる際，顔を主として様々な身体部位へ視線を配ることが報告されている．本論文では，明示的に手や足をタスクに含めることで，全身を含む人物画像において，視線が手や足に集まるかどうかを，最初に実験で確認する．次に，印象単語の間で視線位置分布がどのように変化するかを実験で確認する．

実験協力者に与える具体的なタスクを説明する．印象単語お

よび身体部位を含む以下のタスクを実験協力者へ与え，その回答 (はい，または，いいえ) を得ることとする．

t_1 ：手が美しいか
t_2 ：手に清潔感があるか
t_3 ：手が可愛いか
t_4 ：足が美しいか
t_5 ：足に清潔感があるか
t_6 ：足が可愛いか

次節では，実験協力者の視線位置を計測する手順について述べる．

3.2 計測の手順

実験協力者に与えるタスクの集合を $\mathcal{T} = \{t_1, ..., t_6\}$ で表記する．全ての実験協力者を含む集合を $\mathcal{I} = \{\mathcal{I}_t | t \in \mathcal{T}\}$ とし，$\mathcal{I}_t$ はタスク $t \in \mathcal{T}$ が与えられた実験協力者の集合とする．ある実験協力者が同じ人物画像を 2 回以上見た場合，人物画像の被写体とは初対面であるという本論文の前提条件が崩れてしまう．よって，1 名の実験協力者へは全実験を通して 1 つのタスクのみを与え，同じ人物画像を 2 回見ることは避けた．以下で，視線位置の計測手順を説明する．

P_1 ：タスク $t \in \mathcal{T}$ が与えられた実験協力者の集合 $\mathcal{I}_t$ の中から，1 名の実験協力者 i をランダムに選び出した．なお，手順 P_1 の段階では実験協力者はタスクの内容を知らないこととした．
P_2 ：その実験協力者 i へタスク t について説明し，手順と回答の仕方も説明した．
P_3 ：灰色画像を 2 秒間提示した．
P_4 ：刺激として与える人物画像の集合 $\mathcal{X}$ の中から，重複なしでランダム選択された 1 枚の人物画像を 7 秒間提示した．
P_5 ：黒色画像を 2 秒間提示した．
P_6 ：手順 P_4 で与えた人物画像について，タスク t に対する回答を実験協力者 i から口頭で聞き出した．
P_7 ：手順 P_3 から P_6 を，人物画像の総数 $X = n(\mathcal{X})$ 枚が終了するまで繰り返した．
P_8 ：全ての実験協力者 ($I = n(\mathcal{I})$ 名) が終了するまで手順 P_1 から P_7 を繰り返した．

なお関数 $n(\ )$ は集合の要素数を返すこととする．

4. 視線位置分布の確率表現

4.1 各人物画像における画素注目確率

タスク t を実験協力者 i へ与えた時，ある人物画像を構成する画素へ，視線位置がどれだけ集まるかを表す確率分布について考える．本論文では 1 枚の人物画像を，その画像中の各位置に存在する画素の集合と捉えて $\mathcal{X}_x$ と表記する．ある時刻 (ここではフレーム f と呼ぶ) において，実験協力者 i の視線が，位置 $\boldsymbol{x}_f$ の画素で計測されたとする．本論文では，位置 $\boldsymbol{x}_j \in \mathcal{X}_x$ の画素において，視線がフレーム f で計測される確率分布を式 (1) で表す．

$$p(\boldsymbol{x}_j | t, i, \mathcal{X}_x, f) = \mathcal{N}(\boldsymbol{x}_j | \boldsymbol{x}_f, \boldsymbol{\Sigma}) \tag{1}$$

ただし，$\mathcal{N}(\boldsymbol{x}_j | \boldsymbol{x}_f, \boldsymbol{\Sigma})$ は，平均 $\boldsymbol{x}_f$，共分散行列 $\boldsymbol{\Sigma}$ の二変量正規分布とする．ここで，$p(\boldsymbol{x}_j | t, i, \mathcal{X}_x, f)$ を正規分布で表す理由について述べる．実験協力者は，視線が計測された位置 $\boldsymbol{x}_f$ の画素のみを見ている訳ではなく，その周辺の画素も同時に見ている．これを近似するために正規分布を用いる．なお，$p(\boldsymbol{x}_j | t, i, \mathcal{X}_x, f)$ は式 (2) を満たすとする．

$$\sum_{\boldsymbol{x}_j \in \mathcal{X}_x} p(\boldsymbol{x}_j | t, i, \mathcal{X}_x, f) = 1 \tag{2}$$

提案手法では，共分散行列を $\boldsymbol{\Sigma} = \mathrm{diag}\left(\sigma^2, \sigma^2\right)$ とする．パラメータ σ はカーネルサイズ $k \times k$ に収まるとし，k を式 (3) で設定する．

$$k = \frac{2dh}{l} \tan \frac{\theta}{2} \tag{3}$$

ただし，d はディスプレイから実験協力者の目までの距離，θ は中心窩で見えている角度，l はディプレイの縦の長さ，h はディスプレイの縦の解像度とする．また，それぞれのピクセルは正方形であると仮定する．

視線計測を行う時，瞬きやノイズなどの影響により，視線位置が観測されないフレームがあることに気を付けなければならない．このため，計測毎に総フレーム数が変化していく．以下では，タスク t に取り組む実験協力者 i から，視線が計測されているフレームの集合を $\mathcal{F}_{ti}$ と表記する．式 (1) では，フレーム f を条件に含む時の確率分布を求めた．ここでは，$\mathcal{F}_{ti}$ で周辺化することを考える．実験協力者 i がタスク t を与えられた時，人物画像 $\mathcal{X}_x$ 上の位置 $\boldsymbol{x}_j$ で視線が観測される確率分布を式 (4) で求める．

$$p(\boldsymbol{x}_j | t, i, \mathcal{X}_x) = \sum_{f \in \mathcal{F}_{ti}} p(\boldsymbol{x}_j | t, i, \mathcal{X}_x, f) p(f) \tag{4}$$

総フレーム数を $F_{ti} = n(\mathcal{F}_{ti})$ とし，$p(f)$ を一様分布 $1/F_{ti}$ で近似すると式 (5) となる．

$$p(\boldsymbol{x}_j | t, i, \mathcal{X}_x) = \frac{1}{F_{ti}} \sum_{f \in \mathcal{F}_{ti}} p(\boldsymbol{x}_j | t, i, \mathcal{X}_x, f) \tag{5}$$

なお，$p(\boldsymbol{x}_j | t, i, \mathcal{X}_x)$ は式 (6) を満たすとする．

$$\sum_{\boldsymbol{x}_j \in \mathcal{X}_x} p(\boldsymbol{x}_j | t, i, \mathcal{X}_x) = 1 \tag{6}$$

式 (5) では，個別の実験協力者 i を条件に含む時の確率分布を求めたが，実験協力者の集合で周辺化することを考える．タスク t に取り組む実験協力者の集合 $\mathcal{I}_t$ を用いて，$p(\boldsymbol{x}_j | t, i, \mathcal{X}_x)$ を式 (7) で周辺化する．

$$p(\boldsymbol{x}_j | t, \mathcal{X}_x) = \sum_{i \in \mathcal{I}_t} p(\boldsymbol{x}_j | t, i, \mathcal{X}_x) p(i) \tag{7}$$

ここで，$p(i) = 1/I_t$ の一様分布で近似すると式 (8) となる．

$$p(\boldsymbol{x}_j | t, \mathcal{X}_x) = \frac{1}{I_t} \sum_{i \in \mathcal{I}_t} p(\boldsymbol{x}_j | t, i, \mathcal{X}_x) \tag{8}$$

I_t はタスク t が与えられた実験協力者の総数とする．本論文では，式 (8) の $p(\boldsymbol{x}_j | t, \mathcal{X}_x)$ を画素注目確率と呼ぶ．なお $p(\boldsymbol{x}_j | t, \mathcal{X}_x)$ は式 (9) を満たすとする．

$$\sum_{\boldsymbol{x}_j \in \mathcal{X}_x} p(\boldsymbol{x}_j | t, \mathcal{X}_x) = 1 \tag{9}$$

ここまでに述べた式 (8) の画素注目確率では，同じ人物画像 $\mathcal{X}_x$ のみしか取り扱えないことに注意しなければならない．この理由について以下で述べる．刺激として与える人物画像はア

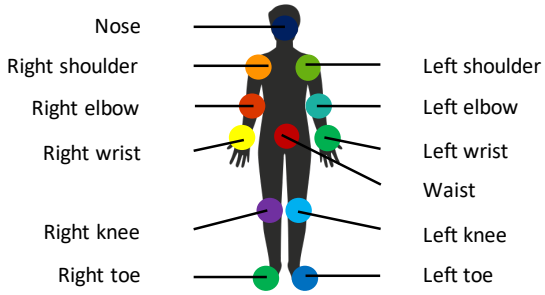

Fig. 2 Body parts for computing the attention probability.

ライメントされていないこと，画像中の被写体の姿勢が異なることが理由として挙げられる．式 (8) の画素注目確率のみでは，異なる人物画像から計測された視線位置分布を，互いに比較することはできない．次節では，様々な人物画像の間で視線位置分布を比較するために，身体部位に着目した確率表現について述べる．

4.2 人物の身体における部位注目確率

画像中に存在する被写体の身体部位へ，観察者の視線位置がどれだけ集まるかを表す確率分布を考える．本論文では身体部位として，顔領域の中心である鼻，胴体領域の左肩と右肩と腰，腕の左関節 2 点と右関節 2 点，足の左関節 2 点と右関節 2 点を取り扱う．それぞれの身体部位を Fig. 2 に示す．合計で 12 個の身体部位を対象とする．前節の式 (8) の $p(\boldsymbol{x}_j|t,\mathcal{X}_x)$ では，タスク t において，人物画像 $\mathcal{X}_x$ が与えられた時に，その画像の位置 $\boldsymbol{x}_j$ で，視線が計測される確率分布を考えた．以下では，視線位置が，身体部位で観測される確率分布を考える．タスク t を与えられた上で，人物画像 $\mathcal{X}_x$ を観察する時，身体部位 b に視線がどれくらい集まるかを表す確率分布を式 (10) で定義する．

$$p(b|t,\mathcal{X}_x)=\sum_{\boldsymbol{x}_j\in\mathcal{X}_x}p(b|\boldsymbol{x}_j,t,\mathcal{X}_x)p(\boldsymbol{x}_j|t,\mathcal{X}_x) \tag{10}$$

なお人物画像 $\mathcal{X}_x$ は，4.1 で述べたように画素の集合である．視線計測の結果から，式 (10) の確率分布を直接求めることは難しい．そこで本論文では，計測された視線位置から身体部位の位置までの距離が近くなるほど，視線位置が身体部位上で計測される確率が高くなると仮定する．提案手法では，$p(b|\boldsymbol{x}_j,t,\mathcal{X}_x)\propto\mathcal{N}(\boldsymbol{x}_j|\boldsymbol{x}_b,\boldsymbol{\Sigma})$ とし，式 (11) を用いることとする．

$$p(b|t,\mathcal{X}_x)=\sum_{\boldsymbol{x}_j\in\mathcal{X}_x}\mathcal{N}(\boldsymbol{x}_j|\boldsymbol{x}_b,\boldsymbol{\Sigma})p(\boldsymbol{x}_j|t,\mathcal{X}_x) \tag{11}$$

ここで，$\boldsymbol{x}_b$ は身体部位 b の中心位置とし，$\mathcal{N}(\boldsymbol{x}_j|\boldsymbol{x}_b,\boldsymbol{\Sigma})$ は平均 $\boldsymbol{x}_b$，共分散行列 $\boldsymbol{\Sigma}$ の二変量正規分布とする．共分散行列は $\boldsymbol{\Sigma}=\mathrm{diag}\left(\sigma^2,\sigma^2\right)$ とし，σ は 4.1 で述べた σ と同じ値とする．なお，$p(b|t,\mathcal{X}_x)$ は式 (12) を満たすとする．

$$\sum_{b\in\mathcal{B}}p(b|t,\mathcal{X}_x)=1 \tag{12}$$

ここで身体部位の集合を $\mathcal{B}=\{b_1,\ldots,b_{12}\}$ で表す．

次に，タスク t が与えられた時，身体部位 b に視線位置がどれくらい集まるかを表す確率分布 $p(b|t)$ を求める．人物画像の集合 $\mathcal{X}$ を用いて，式 (11) の確率分布を式 (13) で周辺化する．

$$p(b|t)=\sum_{\mathcal{X}_x\in\mathcal{X}}p(b|t,\mathcal{X}_x)p(\mathcal{X}_x) \tag{13}$$

ここで，人物画像の総数を X とし，$p(\mathcal{X}_x)$ を一様分布 $1/X$ で近似すると式 (14) となる．

$$p(b|t)=\frac{1}{X}\sum_{\mathcal{X}_x\in\mathcal{X}}p(b|t,\mathcal{X}_x) \tag{14}$$

本論文では $p(b|t)$ を，タスク t が与えられた時の部位注目確率と呼ぶ．

4.3 確率分布間の距離尺度

異なるタスク t を条件とした時の部位注目確率について，それらの分布間の距離がどの程度近いかを推定する尺度について述べる．本論文では，非負性と非退化性を満たす KL(Kullback-Leibler) ダイバージェンス [17] に基づき定義される JS(Jensen-Shannon) ダイバージェンス [18] を用いる．JS ダイバージェンスは，非負性と非退化性に加えて対称性を満たす．具体例として，タスク t_1 の部位注目確率 $p(b|t_1)$，タスク t_2 の部位注目確率 $p(b|t_2)$ の間の距離を考える．$p(b|t_1)$ を p_1，$p(b|t_2)$ を p_2 で表記し，D_{JS} を JS ダイバージェンスとすると，D_{JS} は式 (15) で推定される．

$$D_{JS}(p_1||p_2)=\frac{1}{2}D_{KL}(p_1||p')+\frac{1}{2}D_{KL}(p_2||p') \tag{15}$$

ここで，$p'=(p_1+p_2)/2$ とし，D_{KL} は KL ダイバージェンスとする．D_{KL} は式 (16) で推定される．

$$D_{KL}(p_1||p_2)=\sum_{b\in\mathcal{B}}p(b|t_1)\log\frac{p(b|t_1)}{p(b|t_2)} \tag{16}$$

なお，式 (15) の D_{JS} が小さいほど，二つの確率分布 p_1,p_2 の間の距離が近いと言える．

5. 実験

観察者が身体への印象を判別する際，何処をどれだけ見たかを表す部位注目確率を算出した．まず，5.1 で刺激として与える人物画像について述べ，5.2 で実験条件について述べる．次に，5.3 で部位注目確率を用いてタスク間の比較を行う．最後に，5.4 で人物画像毎の部位注目確率を可視化する．

5.1 刺激として与える人物画像

視線位置分布を求める際に，実験協力者へ刺激として与えた人物画像 $\mathcal{X}_x$ について述べる．本実験で用いた人物画像の一部を Fig. 3 に示す．人物画像の総数を $X=96$ 枚とした．人物画像 1 枚に含まれる被写体の数は 1 名のみとした．被写体の男女比は 1 : 1 とし，その人数は男性 48 名，女性 48 名とした．被写体の頭から足までの全身が画像中に含まれていることとし，その被写体の顔が見えていることとした．被写体の服装として，フォーマルなシーンに相応しい格好，および，カジュアルな格好の 2 種類を用いた．その人数はフォーマル 48 名，カジュアル 48 名とした．フォーマルに加えてカジュアルを含めた理由は比較のためである．なお，被写体が身に着けている装飾品，および，被写体の姿勢には制約を特に設けなかった．被写体は全て成年とした．人物画像中の背景として，被写体以外の他の物体が存在しない背景なし，および，他の物体を含む背景ありの 2 種類を用いた．その枚数は背景なし 48 枚，背景あり 48 枚とした．人物画像サイズを，縦幅が 972 画素となるよ

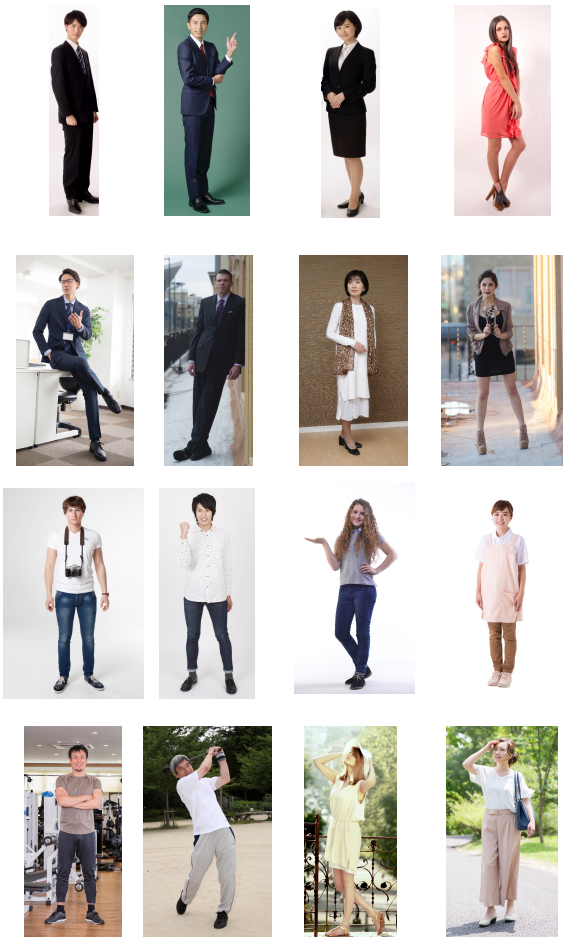

Fig. **3** Examples of person images used in our experiments.

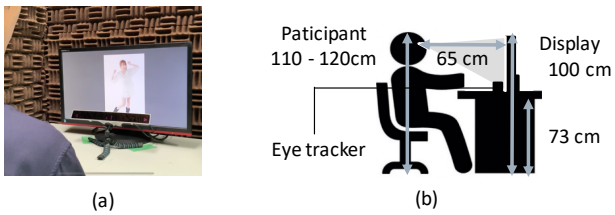

Fig. **4** Locations of the participants and the measurement devices.

Fig. **5** Examples of displaying person images on the display.

うにリスケーリングした．横幅は人物画像毎に異なり，平均は 521 ± 105 画素であった．なお，これらの人物画像は写真素材サイト[*1]からダウンロードした．

5.2 実験条件

実験協力者の総数を $I = 24$ 名 (日本人学生，男性 12 名，女性 12 名) とした．その平均年齢は 22.4 ± 1.0 歳であった．実験協力者へ与えるタスクとして，3.1 で述べた t_1 から t_6 を用いた．各タスクに取り組む人数 I_t を，それぞれで 4 名とした．

[*1] https://www.photo-ac.com/

Fig. **6** Examples of the body parts detected from the person images.

次に，視線計測のセッティングについて述べる．Fig. **4** (a) に計測の様子を，(b) に実験協力者と視線計測装置の位置関係を示す．ディスプレイから 65 センチメートルの位置に実験協力者を座らせた．目の高さが 110 から 120 センチメートルになるよう椅子の高さを実験協力者に調整させた．ディスプレイの大きさは 24 インチ (解像度は 1920 × 1080 画素)，視線計測装置は Gazepoint GP3 HD とした．サンプリングレートは 60 ヘルツとした．なお，視線計測装置の空間分解能の誤差は，約 1 度である (仕様書記載)．センターバイアス [19, 20] を避けるためにディスプレイ上のランダムな位置に人物画像を表示した．ディスプレイ上における人物画像の表示例を Fig. **5** に示す．

提案手法では，4.2 で述べたように，人物画像中の被写体の身体部位が必要となる．これらの身体部位の位置を検出するために，本論文では OpenPose [21] を適用した．検出された身体部位の例を Fig. **6** に示す．刺激として与える人物画像において，身体部位の検出漏れや位置ずれの影響がほぼ無いことを目視で確認した．

式 (3) のカーネルサイズ k は 163 とした．このパラメータを決める際，d は 65 センチメートル，h は 1080 画素，l は 30 センチメートルとした．中心窩の角度が 2 度であること，および，視線計測装置の誤差が約 1 度であることから，θ はマージンを加え 4 度とした．式 (1) の共分散行列 $\mathbf{\Sigma}$ の標準偏差 σ は，$k \times k = 163 \times 163$ のサイズに収まるよう 30 とした．

5.3 部位注目確率を用いた検証

5.3.1 検証の流れ

与えたタスクに含まれる印象単語と身体部位との組み合わせが変化すると，視線位置分布に変化が生じることを実験で検証した．この検証のために，タスク t_1 から t_6 において式 (14) の部位注目確率を算出し，分布間の距離を求めることがまず考えられる．ただし，タスクの組合せ数が多いため，以下の条件に絞る周辺化を行った．

条件 1 ：タスクに含めた身体部位の手，または，足
条件 2 ：タスクに含めた印象単語の美しい，清潔感がある，または，可愛い
条件 3 ：男性実験協力者，または，女性実験協力者
条件 4 ：人物画像中の男性被写体，または，女性被写体

これらの条件を用いて，本実験では以下の流れで検証を進めた．

1. 各条件を満たすように部位注目確率を周辺化した．
2. 周辺化された部位注目確率の間で JS ダイバージェンスを求めた．
3. JS ダイバージェンスの大きさが，条件と条件の間で相対的にどのように変化するかを考察した．

次節で検証方法の詳細と結果について述べる．

Table **1** Conditional probability of body-part-attention for hands or feet.

Body parts	Body-part-attention probability	
	Hands $p_{t_{1,2,3}}$	Feet $p_{t_{4,5,6}}$
Nose (鼻)	16.73	12.73
Right shoulder (右肩)	9.42	3.06
Left shoulder (左肩)	10.42	3.16
Right elbow (右肘)	8.35	2.55
Left elbow (左肘)	9.46	2.24
Right wrist (右手首)	17.00	3.98
Left wrist (左手首)	16.61	4.66
Waist (腰)	7.93	13.54
Right knee (右膝)	1.60	16.04
Left knee (左膝)	1.48	18.90
Right toe (右爪先)	0.49	8.95
Left toe (左爪先)	0.51	10.19

Table **2** Conditional probability of body-part-attention for impression word: beautiful, tidy, or cute.

Body parts	Body-part-attention probability		
	Beautiful $p_{t_{1,4}}$	Tidy $p_{t_{2,5}}$	Cute $p_{t_{3,6}}$
Nose (鼻)	9.46	21.55	13.18
Right shoulder (右肩)	4.99	7.74	5.98
Left shoulder (左肩)	5.42	8.39	6.56
Right elbow (右肘)	4.77	6.00	5.58
Left elbow (左肘)	5.86	5.84	5.85
Right wrist (右手首)	11.44	9.10	10.93
Left wrist (左手首)	11.20	9.07	11.65
Waist (腰)	12.77	10.43	9.00
Right knee (右膝)	11.60	5.96	8.90
Left knee (左膝)	12.85	8.04	9.69
Right toe (右爪先)	4.45	3.86	5.85
Left toe (左爪先)	5.19	4.02	6.83

5.3.2 部位注目確率を周辺化した場合の検証

まず 5.3.1 で述べた条件 1 を満たす周辺化の方法について述べる．手を条件とした時の部位注目確率を，タスク t_1 から t_3 を用いて，$p_{t_{1,2,3}} = (p(b|t_1)+p(b|t_2)+p(b|t_3))/3$ で算出した．タスク t_1 から t_3 には， 3.1 で述べたように，身体部位として手が含まれる．$p(b|t_1)$ から $p(b|t_3)$ は式 (14) により算出した．$p(t)$ は一様分布で $1/3$ とした．同様に，足を条件とした時の部位注目確率を，$p_{t_{4,5,6}} = (p(b|t_4)+p(b|t_5)+p(b|t_6))/3$ で算出した．手，または，足を条件とした時の部位注目確率を Table **1** に示す．表中ではパーセント表記を用いた．まず，$p_{t_{1,2,3}}$ について視線の配り方を確認した．手に属する右手首と左手首では，鼻を除いた他の部位と比べて，注目確率が高かった．顔の中心である鼻の注目確率は，両手首の確率と同程度に高かった．足を構成する部位（両膝，両爪先）の注目確率は，他の部位と比べて非常に低かった．次に，$p_{t_{4,5,6}}$ について視線の配り方を確認した．足を含む下半身を構成する部位（腰，両膝，両爪先）では，鼻を除く上半身（両肩，両肘，両手首）と比べて，注目確率が高かった．鼻の注目確率は，下半身を構成する部位の確率と同程度であった．これらの結果より，身体部位をタスクに含めた場合，その部位に対して，実験協力者の視線が集まることが多いと考えられる．また，顔が明示的にタスクに含まれなくとも，実験協力者は顔へ視線を配ることが多いと考えられる．

次に 5.3.1 で述べた条件 2 を満たす周辺化の方法について述べる．美しいを条件とした時の部位注目確率を，タスク t_1 と t_4 を用いて，$p_{t_{1,4}} = (p(b|t_1) + p(b|t_4))/2$ で算出した．同様に，清潔感があるを条件とした時の部位注目確率 $p_{t_{2,5}}$，可愛いを条件とした時の部位注目確率 $p_{t_{3,6}}$ を算出した．美しい，清潔感がある，または，可愛いを条件とした時の部位注目確率を Table **2** に示す．まず，$p_{t_{1,4}}$ について視線の配り方を確認した．両手首と腰と両膝では，鼻を除いた他の部位と比べて，注目確率が高かった．残りの身体部位の中では鼻の注目確率が高かった．次に，$p_{t_{2,5}}$ について視線の配り方を確認した．鼻の注目確率は，他の身体部位の確率と比べて極端に高かった．最後に，$p_{t_{3,6}}$ について視線の配り方を確認した．鼻の注目確率が最も高く，次に両手首の確率が高かった．これらの結果より，美しいをタスクに含めた条件に絞った場合，両手首と腰と両膝に視線が集まる可能性があると言える．清潔感があるをタスクに含めた条件に絞った場合，顔へ視線が多く集まる可能性が高いと考えられる．可愛いをタスクに含めた条件に絞った場合，顔に視線が集まり，次に両手首に視線が集まる可能性があると考えられる．これらの結果より，タスクに含める印象単語の間で，身体部位に対する視線の配り方に差異が発生する傾向があることを確認した．これらの視線の配り方の差異をさらに定量的に評価するため，次節では，条件と条件の間で JS ダイバージェンスを求め考察する．

5.3.3 条件毎に絞った部位注目確率の間で JS ダイバージェンスを比較検証

(a) ダイバージェンスの算出:

まず，手と足の間で視線の配り方の差異を確認した．手を条件とした時の $p_{t_{1,2,3}}$ と足を条件とした時の $p_{t_{4,5,6}}$ の間の JS ダイバージェンス D_α は以下となった．

$$D_\alpha = D_{JS}(p_{t_{1,2,3}}||p_{t_{4,5,6}}) = 21.18$$

この D_α の値の大きさが持つ意味を， 5.3.1 で挙げた他の条件での JS ダイバージェンスを求めた後に考察する．

次の条件として，印象単語の間で視線の配り方の差異を確認した．美しいを条件とした時の $p_{t_{1,4}}$ と清潔感があるを条件とした時の $p_{t_{2,5}}$ の間の JS ダイバージェンス D_β は以下となった．

$$D_\beta = D_{JS}(p_{t_{1,4}}||p_{t_{2,5}}) = 2.52$$

清潔感があるを条件とした時の $p_{t_{2,5}}$ と可愛いを条件とした時の $p_{t_{3,6}}$ の間の JS ダイバージェンス D_γ は以下となった．

$$D_\gamma = D_{JS}(p_{t_{2,5}}||p_{t_{3,6}}) = 1.25$$

可愛いを条件とした時の $p_{t_{3,6}}$ と美しいを条件とした時の $p_{t_{1,4}}$ の間の JS ダイバージェンス D_δ は以下となった．

$$D_\delta = D_{JS}(p_{t_{3,6}}||p_{t_{1,4}}) = 0.69$$

次に，男性実験協力者と女性実験協力者との間で，視線の配り方の差異を確認した．男性実験協力者の部位注目確率 $p_{M,P}$ と女性実験協力者の部位注目確率 $p_{F,P}$ の間の JS ダイバージェンス D_ϵ は以下となった．

$$D_\epsilon = D_{JS}(p_{M,P}||p_{F,P}) = 0.43$$

最後に，人物画像中の男性被写体と女性被写体との間で，視線の配り方の差異を確認した．男性被写体の部位注目確率 $p_{M,S}$ と女性被写体の部位注目確率 $p_{F,S}$ の間の JS ダイバージェンス D_ζ は以下となった．

$$D_\zeta = D_{JS}(p_{M,S}||p_{F,S}) = 0.20$$

(b) ダイバージェンスの考察:

上記の JS ダイバージェンスの大きさが，条件と条件の間で相対的にどのように変化したかを考察した．手と足の間で求めた D_α は，他の条件で求めた D_β から D_ζ と比べると，大きな差異が見られた．この結果より，タスクに含める身体部位が変わると視線の配り方が大きく変化すると言える．印象単語間で求めた D_β と D_γ は，D_α を除いた D_γ から D_ζ と比べて，差異が大きかった．この結果より，印象単語の一部には，身体部位ほどではないが，性差と比べて，視線の配り方に影響を及ぼすものが存在すると言える．ただし，D_δ は，D_β と D_γ に比べて差異が小さかった．この理由を考察する．辞書[*2]では，美しいという単語の意味の中に，可愛いが含まれていることが述べられている．このことより，両単語の間で部位注目確率が類似したと考えられる．最後に D_ϵ と D_δ からは，実験協力者や被写体の性差による部位注目確率の違いは，身体部位や印象単語の間の違いと比べて小さいことが分かった．

5.3.4 人物画像について他の条件に絞った場合の考察

実験協力者へ刺激として与えた人物画像について， 5.3.1 で述べた条件以外でも絞ることができる．ここでは他の条件を用いた場合について，部位注目確率がどのように変化するかについて評価した．各条件について 5.3.2 と同様に部位注目確率を周辺化し JS ダイバージェンスを求めた．以下に，絞った条件と JS ダイバージェンスを示す．

- 人物画像の背景あり，または，なし $\rightarrow 0.16$
- 被写体の服装がフォーマル，または，カジュアル $\rightarrow 0.17$
- 実験協力者の回答がはい，または，いいえ $\rightarrow 0.32$

前節で求めた D_α から D_δ と比べると，JS ダイバージェンスの値が小さいことが分かった．以上の結果より，人物画像の背景や服装や回答を条件とした場合，身体部位や印象単語を条件とした場合に比べて，部位注目確率に変化が生じにくいと考えられる．

5.4 画素注目確率を用いた比較

ここまでの実験では式 (14) の部位注目確率について評価したが，本実験では式 (8) の画素注目確率について評価した．刺激として与えた人物画像を Fig. **7**(a) に示す．これらの人物画像に対して，各タスクにおいて求めた画素注目確率を Fig. **7**(b) から (g) に示す．図中にて黒色領域は，実験協力者の視線が集まることを表している．逆に，白色領域は視線が集まらないことを表している．

身体部位の手と印象単語の美しいがタスクに含まれる (b) において，手領域に実験協力者の視線が集まっていた．手と清潔感があるがタスクに含まれる (c) において，顔を主とした上半身に視線が配られていた．手と可愛いがタスクに含まれる (d) において，手領域に視線が集まり，顔領域にも視線が配られていた．足と美しいがタスクに含まれる (e) において，足領域に視線が集まっていた．足と清潔感があるがタスクに含まれる (f) において，顔を主として全身に視線が配られていた．足と可愛いがタスクに含まれる (g) において，足領域に視線が集まり，顔領域近くにも視線が僅かに行っていた．これらの結果より，与えたタスクに含まれる印象単語と身体部位の組み合わせが変化すると，視線位置分布に変化が生じるという 5.3 で述べた実験結果が，それぞれの画素注目確率においても見られることが分かった．

5.5 既存の分析研究との共通性の考察

本論文の実験で得た結果と，2.2 で述べた既存の分析研究 [12, 13, 14, 15, 16] の結果との共通性について，以下の各項目について考察した．

実験協力者が取り組むタスク: 既存の分析研究のタスクでは，印象を対象とする点において，本論文のタスクと共通すると考えられる．既存の分析研究では，印象単語を用いて主観評価が実施されたが，単語間の視線の差異について調査されなかった．また，既存の分析研究のタスクでは身体部位を含まないが，本論文のタスクでは身体部位も含めた．

タスクで表示される刺激画像: 既存の分析研究では，各タスクに応じた刺激画像が選択された．被写体が人物であるという点で，本論文の刺激画像と共通すると考えられる．

視線位置を集計する身体部位: 既存の分析研究で用いられた身体部位と，本論文で扱った身体部位との間では，その部位の名称は異なるものの，同じ部位を見ているため共通すると考えられる．ただし，本論文では，胸や腹など胴体を詳細に分ける部位を含めなかった．

実験協力者の内訳: 既存の分析研究では，実験協力者の男女比が，タスクに応じて男性のみ，女性のみ，1 : 1 に設定された．本論文では男女比を 1 : 1 とした．

実験での調査項目: 既存の分析研究では，各身体部位における視線の停留時間が検定された．本論文では，視線の停留点の位置とフレーム数から部位注目確率を算出した．どの身体部位を見ているかを集計するために，視線の停留点を用いる点は共通すると考えられる．

実験結果から得た結論: 既存の分析研究では，与えられたタスクがそれぞれの文献で異なるものの，顔での停留時間が最も長いという同じ結果が報告されていた．本論文では，清潔感があると可愛いの印象単語を含むタスクにおいて，そのタスクに手や足が含まれるものの，顔に視線が集まる確率が最も高いという結果が得られた．この本論文の結果は，既存の分析研究と共通すると考えられる．

以上より，本論文の結論と，既存の分析研究の結論に共通性がある部分が存在すると考えられる．ただし，この共通性の仮説について，検定を用いて実証していないため，その共通性を結論づけることはできない．今後の課題として，本論文の知見と既存の分析研究の知見との共通性を，仮説検証の実験を用いて詳細に分析することが必要である．

6. まとめ

本論文では，人物画像の身体部位に対する印象単語を観察者が判別する際，それぞれの部位に視線位置がどれだけ集まるかを表す確率分布を求める手法について述べた．フォーマルなシーンにおける人物の身体部位について，その印象単語を判別するタスクを観察者に与え，観察者の視線位置を計測した．提案手法の画素注目確率と部位注目確率により，視線位置分布を

[*2] 広辞苑第 7 版

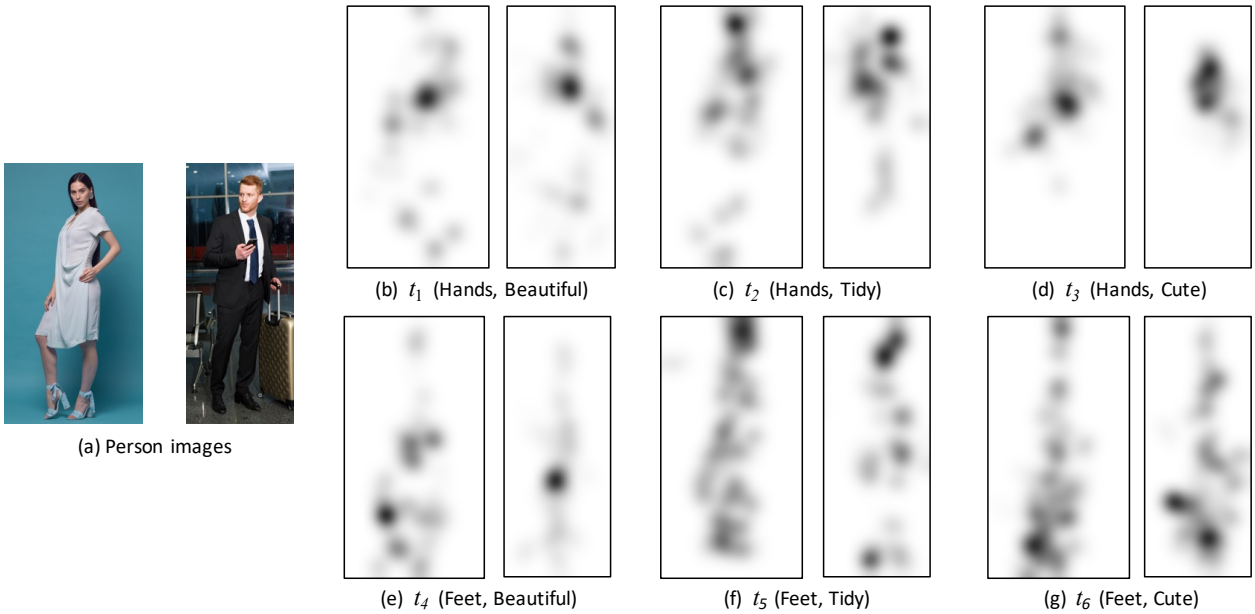

Fig. **7** We show the conditional probability of pixel-attention for each person image and each task. Given the person images (a), the probabilities of pixel-attention (b) to (g) were acquired.

表現した．実験結果より，タスクに含まれる印象単語と身体部位の組み合わせで，注目確率同士の JS ダイバージェンスが変化することから，実験協力者の視線の配り方に差異が生じる可能性があることが明らかになった．

今後の課題として，確率表現された視線位置分布を機械学習や深層学習のアルゴリズムに組み込み，印象推定の精度を評価することが挙げられる．既存手法 [12, 13, 14, 15, 16] は印象推定を目的としていないが，視線位置分布を用いるアイデアは利用できる場合があると考えられるため，精度比較が必要である．また，タスクに含まれる印象単語や身体部位の種類を増やした場合の評価，多様な実験協力者や被写体に対する評価も必要である．実験協力者の人数を増加させた場合の集計を行い，統計的検定を用いて効果の検証を行うことも必要である．

謝　　辞

本研究の一部は，JSPS 科研費 JP20K11864 の助成を受けたものである．

参 考 文 献

1) P. Sudowe, H. Spitzer, and B. Leibe. Person attribute recognition with a jointly-trained holistic cnn model. In *Proceedings of the IEEE International Conference on Computer Vision Workshop*, pp. 329–337, 2015.
2) T. Matsukawa and E. Suzuki. Person re-identification using cnn features learned from combination of attributes. In *Proceedings of the 23rd International Conference on Pattern Recognition*, pp. 2428–2433, 2016.
3) W. Wang, J. Shen, and H. Ling. A deep network solution for attention and aesthetics aware photo cropping. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 41, No. 7, pp. 1531–1544, 2019.
4) X. Zhang, X. Gao, W. Lu, and L. He. A gated peripheral-foveal convolutional neural network for unified image aesthetic prediction. *IEEE Transactions on Multimedia*, Vol. 21, No. 11, pp. 2815–2826, 2019.
5) N. Murrugarra-Llerena and A. Kovashka. Learning attributes from human gaze. In *Proceedings of IEEE Winter Conference on Applications of Computer Vision*, pp. 510–519, 2017.
6) T. Qiao, J. Dong, and D. Xu. Exploring human-like attention supervision in visual question answering. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence*, pp. 7300–7307, 2018.
7) J. Wu, S. Zhong, Z. Ma, S. J. Heinen, and J. Jiang. Gaze aware deep learning model for video summarization. In *Proceedings of the Pacific Rim Conference on Multimedia*, pp. 285–295, 2018.
8) Y. Sugano, Y. Ozaki, H. Kasai, K. Ogaki, and Y. Sato. Image preference estimation with a data-driven approach: A comparative study between gaze and image features. *Journal of Eye Movement Research*, Vol. 7, No. 3, 2014.
9) N. Murrugarra-Llerena and A. Kovashka. Learning attributes from human gaze. In *Proceedings of IEEE Winter Conference on Applications of Computer Vision*, pp. 510–519, 2017.
10) H. Sattar, A. Bulling, and M. Fritz. Predicting the category and attributes of visual search targets using deep gaze pooling. In *Proceedings of IEEE International Conference on Computer Vision Workshops*, pp. 2740–2748, 2017.
11) M. Nishiyama, R. Matsumoto, H. Yoshimura, and Y. Iwai. Extracting discriminative features using task-oriented gaze maps measured from observers for personal attribute classification. *Pattern Recognition Letters*, Vol. 112, pp. 241–248, 2018.
12) B. Philippe, S. J. Gervais, A. M. Holland, and M. D. Dodd. When do people ”check out” male bodies? appearance-focus increases the objectifying gaze toward men. *Psychology of Men and Masculinity*, Vol. 19, No. 3, pp. 484–489, 2018.
13) O. Bareket, N. Shnabel, D. Abeles, S. Gervais, and S. Yuval-Greenberg. Evidence for an association between men’s spontaneous objectifying gazing behavior and their endorsement of objectifying attitudes toward women. *Sex Roles*, pp. 245–256, 2018.
14) B. Dixson, G. Grimshaw, W. Linklater, and A. Dixson. Eye-tracking of men’s preferences for waist-to-hip ratio and breast size of women. *Archives of sexual behavior*, Vol. 40, pp. 43–50, 2009.
15) A. R. Riemer, M. Haikalis, M. R. Franz, M. D. Dodd, D. Dilillo, and S. J. Gervais. Beauty is in the eye of the beer holder: An initial investigation of the effects of alcohol, attractiveness, warmth, and competence on the objectifying gaze in men. *Sex Roles*, Vol. 79, pp. 449–463, 2018.
16) 互恵子, 高田定樹. アイトラッキングによる他者の外見に対する視覚的注意と印象形成の検討. 日本化粧品技術者会誌, Vol. 47, No. 2, pp. 128–134, 2013.
17) *Entropy, Relative Entropy, and Mutual Information*, chapter 2, pp. 13–55. J. Wiley and Sons, Ltd, 2005.
18) B. Fuglede and F. Topsoe. Jensen-shannon divergence and hilbert space embedding. pp. 31–, 2004.
19) M. Bindemann. Scene and screen center bias early eye movements in scene viewing. *Vision Research*, Vol. 50, No. 23, pp. 2577–2587, 2010.
20) G. T. Buswell. *How people look at pictures: A study of the psychology of perception of art*. University of Chicago Press, 1935.
21) Z. Cao, G. Hidalgo, T. Simon, S. Wei, and Y. Sheikh. Openpose: Realtime multi-person 2d pose estimation using part affinity fields. *CoRR, abs/1812.08008*, 2018.