6Q-05

写真撮影スキルの向上を支援する VR トレーニングシステム

小林　大貴†　　長尾　確‡

名古屋大学　大学院情報学研究科†‡

1. はじめに

伝えたいことを簡潔に表現できる写真や画像は情報を「伝達」するためのツールとして用いられるようになった．近年では，スマートフォンと SNS の普及により，写真の撮影と写真の共有が同一の機器で行えるようになった．そのため，写真を共有する流れは加速した．これにより，写真は自分だけが見るものではなく，他の人にも見てもらうものという考え方が浸透した．誰かに見せる以上，写真は「表現」のツールとしても用いられるようになり，写真を上手く撮りたいと願望を抱く人は増加した．

一方で，写真やカメラに関心を寄せる人が抱える課題もある．それは被写体となってくれる人や物を探すことは難しくなかなか練習できないこと，アドバイスしてくれる人が身の周りにいないことなどである．そこで本研究では，写真撮影の練習の場を VR 環境で用意し，利用者が VR 内で撮影した写真に対して評価・フィードバックを行う VR カメラ練習システムを提案する．

2. 写真の評価モデルの構築

本システムの構成を図 1 に示す．本研究では仮想世界内で静止したアバターの撮影を行い，リアルタイムにその写真を評価する．写真評価の仕組みとして，写真の美的を評価するモデル，構図を評価するモデル，色彩を評価するモデルの 3 つを使用する．

2.1 写真の美的評価モデル

美的評価モデルとして，精度と計算コストの観点から NIMA モデル [1]を用いた．NIMA モデルでは 1 枚の画像を入力とし，美的スコアの確率分布を出力する．そして，出力された確率分布の重み付き平均により，美的であるかの 2 値分類を行う．そのため損失関数は，一般的な Binary Cross Entropy ではなく 2 つの確率分布間の距離を測る EMD (Earth Mover's Distance) [2]を用いる．

本研究では，データ収集のため，仮想世界内で 5000 枚の写真を自動撮影した．そして 18 名の評価者により，1 枚の写真に対して 2 名で 5 段階評価のラベル付けをしてもらった．2 名の評価の一致度を表す指標として，Cohen の重み付きカッパ係数 [3]を用いた．本研究では，カッパ係数の値が 0.4 より大きい 3800 枚のデータを使用した．

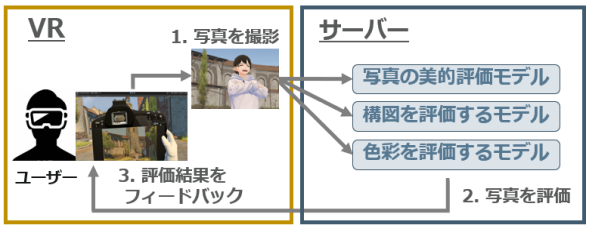

図 1: システムの構成図

図 2: 構図のアノテーションシステム

NIMA モデルの事前学習では写真の大規模データセット [4]を用いて学習を行った．また収集した 3800 枚のデータに対して訓練データとテストデータを 9:1 に分割し Fine-tuning を行った．表 1 に Fine-tuning 前後のテストデータに対する精度を示す．表 1 より，仮想世界内で撮影した写真を用いた Fine-tuning による精度の向上を確認した．

表 1: テストデータに対する精度

	Accuracy	F 値
Fine-tuning 前のモデル	0.605	0.659
Fine-tuning 後のモデル	0.821	0.833

ただし，このモデルでは，写真の何が良くて何が悪かったのか判断できないため，撮影スキルを評価する構図と色彩のモデルを作成する．

2.2 構図を評価するモデル

本研究では図 2 に示すような構図評価のアノテーションシステムを作成した．このシステムでは，同じシーンを撮影した写真を比較することにより，良い画像を選択してもらうという仕様になっている．評価者はカメラ歴 2 年以上の 6 名に協力してもらい，6 名全員に 300 シーンの評価をして頂いた．

また，3 名以上の評価の一致度を表す指標として Fleiss のカッパ係数 [5]を用いた．本研究では，カッパ係数が 0.4 を超える 5 名の組み合わせを用いて，優劣のついたペア画像を 13000 組作成した．

構図を評価するモデルとして VEN モデル [6]を

VR Training System for Improving Photography Skills
†KOBAYASHI, Hiroki (kobayashi@nagao.nuie.nagoya-u.ac.jp)
‡NAGAO, Katashi (nagao@i.nagoya-u.ac.jp)
†‡Graduate School of Informatics, Nagoya University

用いた．VEN モデルでは，学習時における入力は優劣のついたペア画像(画像I_i, 画像I_j)とし，出力は入力画像に対する構図のスコア$(f(I_i), f(I_j))$とする．また損失関数は(1)式のように定義される(画像I_iの方が画像I_jより構図が優れている場合)．

$$loss(I_i, I_j) = max\left(1 + f(I_j) - f(I_i)\right) \quad (1)$$

本研究では，構図の大規模データセット(CPC データセット) [6]を用いて事前学習を行った．そして，作成した 13000 組のペア画像を 9:1 に分割して Fine-tuning を行った． FLMS データセット [7]を用いてモデルの評価を行った結果を表 2 に示す．評価指標は，先行研究 [6]に基づき IOU と Displacement Error を用いた．IOU とは予測領域と正解領域の和集合のうち 2 つの領域が重なっている割合を表す指標である．また Displacement Error とは正規化された予測領域と正解領域において各辺の差分を平均した指標である．

表 2: FLMS データセットに対する精度

	IOU ↑	Disp.Error ↓
先行研究(VEN)	**0.8365**	**0.041**
作成したモデル	**0.8531**	**0.033**

2.3 色彩を評価するモデル

構図と同様に，色彩に関してもデータセットを作成した．色彩評価では 500 枚の写真に対してコントラスト，色の調和，明るさの 3 項目を 2 段階評価してもらい，総合的な色の観点で 5 段階評価してもらった．評価者は構図と同様である．

本研究では文献 [8]に基づいて色彩調和モデルを作成した．このモデルでは，入力画像(224×224×3)を均等に分割してパッチ画像(16×16×3)を生成する．そして中心パッチ(16×16×3)と 4 近傍の隣接パッチ(16×16×3)を逐次的に入力する．中心パッチだけではなく隣接パッチも入力することで，パッチ間の色の調和を学習する．またモデルの出力は入力画像の調和確率であり，損失関数は Binary Cross Entropy を使用した．本研究では 500 枚の写真データを 8:1:1 に分割し，訓練データについては画像を回転させることによりデータ拡張を行った．表 3 にモデルの精度を示す．

表 3: validation と test データに対する精度

	Accuracy	F 値
validation データ	**0.780**	**0.820**
test データ	**0.780**	**0.820**

3. 構図を推薦するモデル

写真を評価するだけではなく，撮影に関してアドバイスしてくれる機能があると望ましい．そこで本研究では，VPN モデル [6]に基づいて構図の推薦機能を実装する．

VPN モデルの入力は 1 枚の画像 I であり，出力は 500 個の構図のスコア$(g(I_1), \dots, g(I_{500}))$とした．また，正解データは VEN モデルを用いて作成する．具体的には，VEN モデルによって，事前に定義した 500 個のアンカーボックス(入力画像をトリミングする領域)に対応する構図のスコアを出力し，それを VPN モデルの正解データとする．

本研究では， CPC データセット [6]を用いて事前学習を行った．そして，仮想世界内で撮影した 2000 枚の縦画像(2:3)と 2000 枚の横画像(3:2)を用いて正解データを作成し Fine-tuning を行った．FLMS データセット [7]を用いてモデルの評価を行った結果を表 4 に示す．

表 4: FLMS データセットに対する精度

	IOU ↑	Disp.Error ↓
先行研究(VPN)	**0.8352**	**0.044**
作成したモデル(縦)	**0.8573**	**0.032**
作成したモデル(横)	**0.8604**	**0.032**

4. まとめと今後の課題

本研究では，仮想世界内においてアバターの写真を撮影し，評価・フィードバックを行うことで，写真撮影スキルの向上を支援する VR Photo トレーニングシステムを提案した．

今後の課題としては，ユーザーに合わせたアドバイス生成が挙げられる．モデルがユーザーの特性を理解した上でどこをどう直せば良いかについてフィードバックする仕組みである．

そのためのアプローチとして，例えば構図評価をさらに分割し，複数の細分化モデルを作成することにより，評価結果をレーダーチャートで表示することなどが考えられる．さらに，ユーザーの活動を記録し，苦手項目については，具体的に良い例を画像で示しながらアドバイスするなどの仕組みによって，撮影スキルのさらなる向上支援が可能になると思われる．

参考文献

[1] H. Talebi and P. Milanfar, Nima: Neural Image Assessment, IEEE Transactions on Image Processing, 27(8), 3998-4011, 2018.

[2] L. Hou, et al., Squared Earth Mover's Distance Based Loss for Training Deep Neural Networks, arXiv preprint arXiv:1611.05916, 2016.

[3] J. Cohen, Weighted Kappa: Nominal Scale Agreement with Provision for Scaled Disagreement of Partial Credit, Psychological Bulletin, 70(4), 213-220, 1968.

[4] N. Murray, et al., AVA: A Large-Scale Database for Aesthetic Visual Analysis, In CVPR, 2012.

[5] J. R. Landis and G. G. Koch, The Measurement of Observer Agreement for Categorical Data, Biometrics, 33, 159-174, 1977.

[6] Z. Wei, et al., Good View Hunting: Learning Photo Composition from Dense View Pairs, In CVPR, 5437-5446, 2018.

[7] C. Fang, et al., Automatic Image Cropping Using Visual Composition, Boundary Simplicity and Content Preservation Models, In ACM Multimedia, 2014.

[8] P. Lu, et al., Gated CNN for Visual Quality Assessment Based on Color Perception, Signal Process Image Comm., 72, 2018.