

[研究助成 (B)]

和菓子 IT アーカイブと利活用

—— 質感・外観の記録復元技術の構築と新規和菓子創成支援 ——

Wagashi IT Archiving and Utilization

—— Technology for Recording and Reconstructing Texture and Appearance and Supporting the Creation of New Japanese Sweets ——

2221903



研究代表者

東京都市大学 メディア情報学部
情報システム学科

准教授

中 村 裕 美

[研究の目的]

本研究の目的は、無形文化財の記録伝承や発展を人間と機械が調和して推進するために、材質的に記録・保存の難しい和菓子の外観・質感等のデータを HCI 技術を用いて再利用可能な形でアーカイブすることである。また、それら外観・質感のアーカイブ情報に和菓子に関する各種情報を紐づけ、人間の発想や指示を基とした AI による新たな和菓子生成を行い、弥生時代より連綿と紡がれ現代まで進化した和菓子という伝統の継承を助けるだけでなく人間と機械の調和による新たな和菓子文化を構築することである。

[研究の内容、成果]

1. ロボットアームと三次元再構成技術 NeRF を用いた和菓子の外観保存手法の構築

日本の無形文化遺産である和食、そしてその一角をなす和菓子は基本的に長期的な保存を行うことが難しい。一方、和菓子は季節や風土、文化が反映された装飾的な食品であり、見た目を含めて使用用途が定まる。また茶会のプログラムに相当する茶会記に菓子に関する記述が記録されたことで、後世で閲覧者は茶会で使用された菓子の推移等を読み解くことができる。茶会記は一部図を含む文章で記録されているが、

外観的な特徴を実物と近い 3 次元形状および質感を有した形でアーカイブできれば、文化の保全、技術継承を行う人材への教育に活用できるだけでなく、新たな発見や理解の一助となる。

外観や形状の 3 次元アーカイブは、レーザースキャンやフォトグラメトリ等がこれまで用いられてきた。これらでアーカイブされる対象は、建築物や美術品などが主である。食分野においては食事量の推定や収量予測のための体積推定に用いられているが、食品の外観のアーカイブ目的では活用されていないのが現状であった。また、既存手法では表面輝度が一定となり、いかなる物体でも石膏のような質感で表現されてしまうこと、これらを改善するには画質とフレームレートの上にトレードオフが生じることなどが問題であったが、近年提案された Neural Radiance Fields (NeRF) では、ニューラルネット上で放射輝度と密度情報を学習させることにより、光沢感や透明感など複雑な質感を再現できる。加えて学習データに含まれない新奇な未知の視点にも対応したレンダリングを可能にしているため、本研究で対象とする和菓子が有する光沢感や透明感の再現に適していると考え、NeRF による三次元再構成を対象とした。

NeRF を活用し高精細な三次元再構成を行うには、被写体を中心に周囲を過不足なく撮影した画像群が必要となる。特に被写体の一部が特

殊な外観（光沢など）や凹凸の多い複雑な形状を有する場合、その箇所はより多くの画像データを必要とする。画像の不足やブレなどは再構成結果にひずみや不自然な浮遊物や曇り、ゴースト（視点によって同一物体が複数表示される現象）を生じさせる原因となりえる。

このような点を考慮して画像データを収集することは多くのユーザにとって簡易な作業とは言い難い。既存アプリケーションではオブジェクトの外周を撮影した動画から画像を切り出す機能が付随したものもあるが、機械的にフレームを抽出するため、適した画像が抽出されとは限らない。そして画像データ（最低でも 30 枚程度は必要とされる）を一枚ずつ撮影する場合、ユーザは撮影済みの視点と枚数を記憶する必要がある。これも一部のアプリケーションでは撮影済みか否かを可視化するなどの工夫が行われているが、特定範囲の画角を集中的に撮影させるような支援は行われていない。

さらには、学習データへの画像の追加は精度の向上に寄与し得るため、三次元再構成結果を参考に追加で撮影することも推奨されているが、精度の向上に寄与する撮影ができるかはユーザの経験や知識に依存する。このような撮影や三次元再構成において生ずる問題を解決できれば、専門家でなくともこれら技術を活用しやすくなると考えられる。

本研究では、専門的知識を有さずとも利用しやすい三次元再構成システムの構築、及び三次元再構成結果の精度向上のためのシステムとして、ロボットアームを用いた撮影システムと 2 つの再撮影システムを提案する。再撮影システムは、ユーザフィードバックや PSNR 値により、三次元再構成結果の改善に適した撮影ポイントを特定し、再撮影を行う。また、既存の手動撮影と比較を行い、自動撮影システムと再撮影システムの性能を検証する。

1.1 ロボットアームによる自動撮影システム

ロボットアームによる撮影システムは、三次元再構成対象となる被写体の画像データを収集

するためのシステムである。当システムの利用によって、ユーザは指定した和菓子を 3 次元再構成するのに必要とされる画像の収集を、自身が撮影せずとも取得することができる。また、ロボットアームでの自動撮影は、人が撮影するのに比べ手ブレ等を軽減できること、後述する再撮影位置の正確な指定ができることから、3 次元再構成の精度向上にも寄与できると考えている。

ロボットアーム自動撮影システムは、ロボットアームとして UFACTORY (株) xArm7、撮影用カメラとして Apple iPhone15 Pro を用いて構築している。ロボットアームの制御は Unity ROS Tool にて実装している。これは Unity の ROS TCP Connector 上に構築され、Unity の GUI を通じてロボットアームの制御システムと通信できるものである。ここで撮影可能な被写体は、ロボットアームで周囲を撮影可能な約 10 cm 四方の立方体に収まる静止したオブジェクトまでとなり、本研究の対象となる和菓子は十分その領域内に収まる。

取得画像群を用いた 3 次元再構成には、NeRF を高速化した手法の一つである instant-ngp を用いた。NeRF では 3 次元再構成時の学習のために画像とともにカメラ姿勢行列を必要とする。そのため、撮影した画像からカメラ姿勢行列を求める前処理が必要である。NeRF 及び instant-ngp では通常 COLMAP 等の SfM (Structure-from-Motion) 手法を用いて画像群からカメラ姿勢を推定するが、この自動推定は誤差が生じやすく、高精細な 3 次元再構成には向かない。そのため、本研究では撮影対象の下に ChArUco マーカと呼ばれるマーカーボードを配置し、キャリブレーション及び座標系の変換を行った上で、カメラ位置を取得している。この処理は、将来的には Hand-Eye Calibration によるロボットの手先姿勢とカメラのキャリブレーションを行うことで代替できるため、図 1 のように和菓子とともにマーカーボードが映ることなく撮影及び再構成が可能に

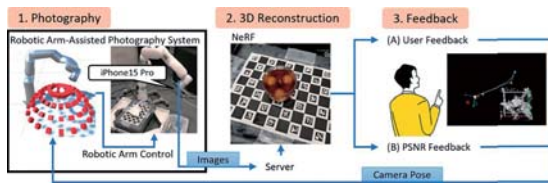


図1 本研究で提案するシステムの概念図

なると考えられる。

ロボットアームによる自動撮影は、被写体を中心とした半球状に設定された撮影ポイント（図1左 赤ボックス参照）で行われる。この際、ロボットアームは赤ボックスの座標で一時停止し、先端のカメラが被写体の画像を撮影した後に次の座標まで移動、これをすべて指定座標で撮影が終了するまで繰り返す。

1.2 フィードバックによる再撮影システム

1.1にて構築したロボットアームによる自動撮影システムによって、被写体周囲の半球を網羅的に撮影することが可能となるため、3次元再構成結果も一定の精度が得られるように思われる。しかし、被写体の質感によっては、均一な撮影だけでは十分な精度を得られない場合がある。特に光沢や凹凸、透明感を持つ部分が有る被写体、一部に細かい質感を持つ被写体などは、その箇所を重点的に再撮影しデータに加えることで、再構成の精度が向上する。そこで本研究ではユーザによるフィードバックを用いた再撮影システムと、画質を基に再撮影候補を選出する再撮影システムの2つを構築した。

1.2.1 ユーザフィードバックによる再撮影システム

ユーザフィードバックによる再撮影システムは、3次元再構成結果に対し、ユーザの必要性に応じた改善要求を反映させられる点がメリットである。例えば、銘の焼印が押された部分をより詳細に閲覧できるようにしたい、といった希望などをコンピュータで推測させるのは難しい。一方、ユーザはUI上で3次元再構成結果を見て精度の低い視点を発見することは比較的容易だが、UI上の表示を基に再撮影すべき実空間上の位置を割り出すのは難しいと考えられ

る。そのため、本システムでは再撮影も1.1のロボットアーム撮影システムを用いて行う。

ユーザフィードバックによる再撮影システムでは、ユーザはGUI上で3次元再構成結果の確認と、再撮影ポイントの指示を行う。ユーザは3次元再構成結果を視点を変えながら閲覧し、曇りやゴースト等が生じている視点や、より詳細に撮影したい視点を発見した際にGUI上に配置された“Add from Cam”ボタンを押下する。このボタンは本来のinstant-ngp GUI上でも用意されているもので、3次元再構成結果をユーザが指定したカメラパスで動画出力するためのものである。本研究ではこの機能で取得できるカメラポーズ情報をロボットアームでの再撮影時のアーム先端位置（カメラ据付位置）情報として用いるために、JSON形式で取得できるよう変更し、OpenGL座標系からNeRF座標系（NeRFでの再学習に使用）、Unity座標系（ロボットアームの制御に使用）へ変換している。

1.2.2 システムフィードバックによる再撮影システム

1.2.1で紹介したユーザフィードバックシステムは、ユーザの改善要求を反映し易い点はメリットで有る一方、すべての再撮影ポイントを人が指定するのは非効率でもある。特に、曇りやゴースト等が発生している画角を再撮影すべきであることは、ユーザからのフィードバックがなくても判断可能で、このような再撮影候補を算出し再撮影するシステムを構築することで、ユーザにかかる負担を減らすことができると考えられる。

本研究では、画質の客観評価手法として用いられるPeak Signal-to-Noise Ratio (PSNR)を用い、事前に指定した30箇所の評価用座標位置で再構成結果と画像とのPSNR値を求め、PSNR値が低い撮影ポイント10箇所を再撮影候補として選出するシステムを構築した。この評価用座標位置は1.1で述べた自動撮影用の座標位置とは異なる座標であり、事前に評価用の

画像データセットとして、評価用座標位置での実物体の写真を撮影して比較する必要がある。

1.3 評価

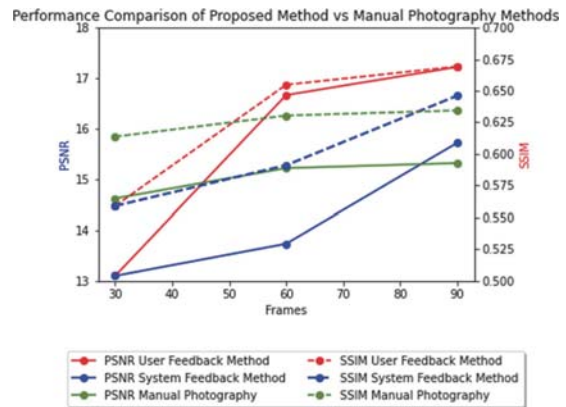
1.1, 1.2で構築したロボットアームによる自動撮影システム及び2つの再撮影システムが和菓子の外観記録復元技術として機能するか、特に再撮影システムによって、3次元再構成結果を改善できるかを検証するために、従来の手動での撮影手法との比較を行った。評価では、表1のとおり、30, 60, 90枚で3次元再構成を行い、NeRFの画質評価で一般的に用いられるPSNR値とStructural Similarity Index (SSIM値)を算出した。

表1 評価に用いた画像データセットの詳細

	30枚	60枚	90枚
手動撮影	手動撮影 30枚	手動撮影 60枚	手動撮影 90枚
ユーザ フィード バック	自動撮影 30枚	自動撮影30枚＋ ユーザフィード バックでの再撮影 30枚	自動撮影30枚＋ ユーザフィード バックでの再撮影 60枚
システム フィード バック	自動撮影 30枚	自動撮影30枚＋ システムフィード バックでの再撮影 30枚	自動撮影30枚＋ システムフィード バックでの再撮影 60枚

手動撮影手法は被写体の周囲を撮影者が移動しながら撮影した動画から50フレームごとに静止画を抽出したものをを用いた。ユーザフィードバック・システムフィードバックにおいては、双方ロボットアームによる自動撮影30枚をベースとして用い、データセット数60枚、90枚での評価においては各々のフィードバックシステムで再撮影した画像を加えたもので実施している。画像サイズはすべて1920×1080ピクセルで統一している。

各手法でのPSNR値とSSIM値は図2のとおりである。手動撮影手法(30枚)とロボットアームによる自動撮影のみ(30枚)を比較すると、手動撮影手法のほうがPSNR・SSIM双方とも高い。一方、フィードバックによる再撮影が加わることによって、90枚使用時には手動撮影手法に比べユーザフィードバック手法、システムフィードバック手法のPSNR・SSIM



実線：PSNR値、点線：SSIM値、赤色：ユーザフィードバック手法、青色：システムフィードバック、緑色：手動撮影手法

図2 各条件での三次元再構成結果における画質評価の結果

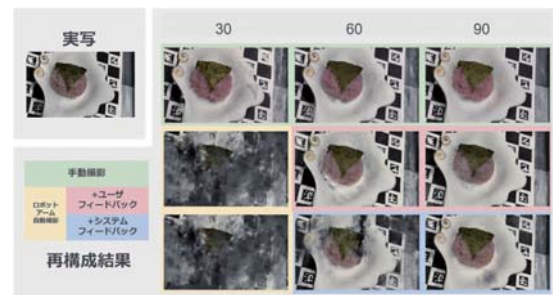


図3 各条件での三次元再構成結果の上部俯瞰視点

値がより高い結果、すなわち再構成結果が鮮明であるという結果が得られた。ユーザフィードバック手法においては、データセットが60枚の時点で手動撮影手法を上回る結果となっている。

図3からも、ロボットアームでの自動撮影のみでは曇りが多く出ているのに対し、フィードバックで改善されている様子が見られる。ユーザフィードバックとシステムフィードバックの精度差の理由としては、システムフィードバックが指定されたポイントの中から再撮影が必要なポイントを探索するのに対し、ユーザフィードバックはより広範な視点から再撮影が必要な箇所を指定できたためと考察している。

本研究では、ロボットアームによる自動撮影とユーザフィードバックによる再撮影を組み合わせた撮影システムによって既存手法に比べ精細な三次元再構成が可能となることが示された。この3次元再構成結果を裸眼立体視ディスプレイ

イで表示することで、より実在感のある和菓子アーカイブの閲覧も可能となる。

2. 和菓子の新規創造を助ける AI 技術の構築

本研究では現存の和菓子の外観・質感を保存し、後世の参考資料として役立てるだけでなく、新しい和菓子のデザイン考案の支援や、そのための技術構築も目的としている。本研究の採択と同時期より、深層学習全体の進歩を受けてテキストからの 2D, 3D 生成や画像からの 3 次元物体の生成等、生成系 AI 技術が急激な進化を見せている。既存の大規模データセットを用いる事で一般的な形状の物体は生成が容易となる一方、和菓子のような特定文化圏で作成されたものの形状の生成に対してはこれら技術が十分に活用可能かは未知である。そのため、現在の Text to Image, Illust to Image, Text to 3D 技術を用いた新たな和菓子デザインに関する調査研究を行った。また、既存の和菓子の種類と傾向は材料や季節をベースとした分類がなされ、まとめられた書籍等も存在するが、それら分類同士の関係を踏まえ俯瞰的に把握することで、新しい和菓子のジャンルを創り出すことも可能だと考えられる。そのため、既存の和菓子情報をもとに画像情報とテキスト情報を用いたクラスタリングを実施し、俯瞰的な把握に寄与できるか検討した。

図 4 はエッジ形状を参考に既存和菓子画像とシンプルな図形図案をもとに ControlNet を用いて生成した和菓子イメージと、それらイメージから ProlificDreamer を用いて 3D 形状を生成したものである。技術的には、3D 形状の生成において裏面へのデザイン付与が行われる等、通常の和菓子ではデザインされにくいと考えられるような生成結果となっているため、ファインチューニングが必要であると考えられる。また、生成されたデザインにおいては、和菓子のような見た目ではあるものの、和菓子のデザインの根底にある季節や風土の反映までは至っていない。そのため、当該手法で得られた結果を



一段目左：参照元の和菓子画像
右：左の画像からエッジを検出し、生成した和菓子イメージ
二段目：一段目で生成した和菓子イメージから 3D 形状を生成したもの
三段目：シンプルな図形図案から生成した和菓子イメージ
四段目：三段目で生成した和菓子イメージから 3D 形状を生成したもの

図 4 既存の生成系技術を用いて 2D, 3D 生成を行った例



図 5 文章・画像情報を用いた和菓子のクラスタリング

閲覧しながら、和菓子製作者が修正点やアイデアなどをあげ、それを再反映させるシステムが必要と考えられる。

図 5 は t-SNE を用いたクラスタリング結果である。煎餅類、飴類、餅類が各々近い位置にあることから大まかな分類は当手法で可能だと考えられる。一方本検討時では各地方での同種の和菓子の比較などは行えていないため、データの拡充と再検討が望まれる。

[成果の発表, 論文など]

堀部咲歩, 中村裕美, 廣井裕一, Émilie Fabre, 義平真規, 暦本純一. 食品の三次元外観アーカイブ構築に向けた自動撮影システムの検討, 信学技報, vol. 123, no. 433, MVE2023-62, pp. 114-119, 2024.