FoodChangeLens: CNN-based Food Transformation on HoloLens

Shu Naritomi, Ryosuke Tanno, Takumi Ege, Keiji Yanai
The University of Electro-Communications, Tokyo, Japan
Introduction: HoloLens

- HoloLens is a device that realizes Mixed Reality which Microsoft develops.

- When user wearing HoloLens, the hologram becomes visible in the real space, various functions are incorporated to realize that.
Introduction: Convolutional Neural Network (CNN)

Pix2Pix
[Berkeley CVPR 2017]
- learning a pair of line drawing and color image.
- Automatic coloring from line image

You Look Only Once
[Joseph Redmon CVPR 2016]
- Detection of position of object and classification
Introduction: Convolutional Neural Network (CNN)

 대하여: Food image translation

◇ learning some category food image.
◇ Convert different category food image.
Objective

Deploy image generation technology on HoloLens and convert food categories in the real space.

VIDEO
Application overview

Real world

Rendering

Take photo

Server

Conditional CycleGAN

Convert Image

U-Net

segmentation

Image + category

HoloLens

Convert to projected coordinates

Projection Matrix

Convert to Camera coordinates

Camera Matrix

Convert to UV coordinates

And

Associate all vertices with texture

Vertices in world coordinates

Spatial mapping meshes

Geometry shader

Triangular polygon

Texture UV coordinate

Geometry shader

Triangular polygon

Texture UV coordinate

Image

Converted Texture

Converted Texture

Spatial mapping meshes

HoloLens
Application overview

Take photo

Rendering

Real world

Server

Conditional CycleGAN

Convert Image

U-Net

segmentation

Image + category

Converted Texture

Vertices in world coordinates

Spatial mapping meshes

HoloLens

Convert to UV coordinates

And Associate all vertices with texture

Geometry shader

Triangle polygon

Texture UV coordinate

image

Convert to Camera coordinates

Camera Matrix

Convert to projected coordinates

projection Matrix

Spatial mapping meshes
Application overview

Real world

Rendering

Take photo

Image + category

Server

Convert Image

Conditional CycleGAN

U-Net

segmentation

HoloLens

Convert to Camera coordinates

Camera Matrix

Convert to Camera coordinates

Camera Matrix

Convert to projected coordinates

projection Matrix

Convert to UV coordinates

Geometry shader

Texture coordinate

Triangular polygon

Vertices in world coordinates

Spatial mapping meshes

Convert to UV coordinates

And

Associate all vertices with texture

Converted Texture

image
Application overview

Take photo

Rendering

Real world

Image

Convert to projected coordinates

projection Matrix

Convert to Camera coordinates

Camera Matrix

Geometry shader

Triangular polygon

Texture UV coordinate

Vertices in world coordinates

Spatial mapping meshes

Convert to UV coordinates

And Associate all vertices with texture

Converted Texture

Server

Conditional CycleGAN

U-Net

segmentation

Convert Image

Image + category

HoloLens
Application overview

Real world

Take photo

Rendering

Convert to projected coordinates
Projection Matrix

Convert to Camera coordinates
Camera Matrix

Convert to UV coordinates
Vertices in world coordinates
Spatial mapping meshes

Geometry shader
Triangular polygon
Texture UV coordinate

Converted Texture

Image + category

Conditional CycleGAN
Convert Image

U-Net
segmentation

Server

HoloLens
Application overview

Take photo

Rendering

Server

Real world

Convert to Camera coordinates

Camera Matrix

Convert to projected coordinates

projection Matrix

Geometry shader

Triangular polygon

Texture UV coordinate

Convert to UV coordinates

Vertices in world coordinates

Spatial mapping meshes

Convert to Camera coordinates

Camera Matrix

Convert to UV coordinates

And

Associate all vertices with texture

Converted Texture

Image + category

Image

Conditional CycleGAN

Convert Image

U-Net

segmentation
Image conversion system

Convert to projected coordinates
- projection Matrix

Convert to Camera coordinates
- Camera Matrix

Spatial mapping meshes

Convert to UV coordinates
- Vertices in world coordinates
  - And Associate all vertices with texture

Geometry shader
- Triangular polygon
  - Texture UV coordinate

Converted Texture

Image + category

Real world

Rendering

Take photo

Server

Conditional CycleGAN
- Convert Image

U-Net
- segmentation
Image conversion system: Conditional CycleGAN

1. Convert with high quality
2. Convert to multiple food category by one network
### Image conversion system : Dataset

<table>
<thead>
<tr>
<th>Category</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ramen</td>
<td>74,007</td>
</tr>
<tr>
<td>Curry rice</td>
<td>34,216</td>
</tr>
<tr>
<td>Fried rice</td>
<td>27,854</td>
</tr>
<tr>
<td>Fried noodle</td>
<td>24,760</td>
</tr>
<tr>
<td>Rice</td>
<td>21,324</td>
</tr>
<tr>
<td>Beef bowl</td>
<td>18,396</td>
</tr>
<tr>
<td>Chilled noodle</td>
<td>13,499</td>
</tr>
<tr>
<td>Meat spaghetti</td>
<td>7,138</td>
</tr>
<tr>
<td>Eel bowl</td>
<td>5,329</td>
</tr>
<tr>
<td>Buckwheat noodle</td>
<td>3,530</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>230,053</strong></td>
</tr>
</tbody>
</table>

Total amount : 230k images
HoloLens application

1. Take photo
2. Rendering
3. Convert to Camera coordinates
4. Camera Matrix
5. Triangular polygon
6. Texture UV coordinate
7. Geometry shader
8. Vertices in world coordinates
9. Spatial mapping meshes
10. Convert to UV coordinates
11. And Associate all vertices with texture
12. Converted Texture
13. Image + category
14. Image to Camera coordinates
15. Projection Matrix
16. Conditional CycleGAN
17. U-Net
18. segmentation
19. Real world

Server
HoloLens application : main used functions

- HoloLens has an API that can acquire
  - camera matrices
  - projection matrices.

- Spatial mapping (Spatial recognition)
  - HoloLens uses spatial recognition using cameras, depth cameras, etc., and creates 3d meshes according to the shape of real objects.
HoloLens application: System

1. Take photo
2. Rendering
3. Real world

- Convert to projected coordinates (projection Matrix)
- Convert to Camera coordinates (Camera Matrix)
- Vertices in world coordinates
- Spatial mapping meshes

- Geometry shader
  - Triangular polygon
  - Texture UV coordinate

- Convert to UV coordinates
  - And Associate all vertices with texture

- Converted Texture

4. Server
- Image + category
- Convert Image
- Conditional CycleGAN
- U-Net
- segmentation

5. HoloLens
HoloLens application: System

- **Take photo**
- **Server**

**HoloLens**

- **Convert to projected coordinates**
  - Projection Matrix

- **Convert to Camera coordinates**
  - Camera Matrix

- **Convert to UV coordinates**
  - Texture UV coordinate
  - Sprite polygon

- **Vertices in world coordinates**

- **Spatial mapping meshes**

- **Geometry shader**
- **Image**
- **Real world**

- **Rendering**
- **Take photo**

- **Converted Texture**

- **Conditional CycleGAN**
- **U-Net**
- **Segmentation**

- **Convert Image**

- **Image + category**
HoloLens application: System

Convert to projected coordinates
Projection Matrix

Convert to Camera coordinates
Camera Matrix

Convert to UV coordinates
Texture UV coordinate

Geometry shader
Triangular polygon

Image

Converted Texture

Vertices in world coordinates
Spatial mapping meshes

Rendering

Take photo

Server

Image + category
Conditional CycleGAN
Convert Image
U-Net
segmentation
HoloLens application: System

- **Take photo**
- **Rendering**
- **Real world**
- **Server**
- **Image + category**
- **U-Net**
- **Conditional CycleGAN**
- **Convert Image**
- **segmentation**

**Workflow:**
1. **Spatial mapping meshes**
2. **Convert to Camera coordinates**
   - Camera Matrix
3. **Convert to projected coordinates**
   - Projection Matrix
4. **Convert to UV coordinates**
   - Triangular polygon
   - Texture UV coordinate
5. **Convert to UV coordinates**
   - Vertices in world coordinates
   - Associate all vertices with texture
6. **Converted Texture**
7. **Image**
8. **Geometry shader**
   - Texture UV coordinate
9. **Real world**

**Key Elements:**
- **HoloLens**
- **Camera Matrix**
- **Projection Matrix**
- **Geometry shader**
- **Triangular polygon**
- **Texture UV coordinate**
- **Vertices in world coordinates**
- **Spatial mapping meshes**

**Server Tasks:**
- **Conditional CycleGAN**
- **U-Net**
- **segmentation**
HoloLens application: System

1. Take photo
2. Rendering
3. Server
4. HoloLens
5. Image + category
6. Convert Image
7. Conditional CycleGAN
8. U-Net segmentation
9. Converted Texture
10. Converted Texture (meshes)
11. Spatial mapping meshes
12. Vertices in world coordinates
13. Geometry shader
14. Texture UV coordinate
15. Triangular polygon
16. Projection Matrix
17. Camera Matrix
HoloLens application: System

Real world

- Rendering
- Take photo

Convert to projected coordinates
- projection Matrix

Convert to Camera coordinates
- Camera Matrix

Vertices in world coordinates

Spatial mapping meshes

Geometry shader
- Triangular polygon
- Texture UV coordinate

Convert to UV coordinates
- And Associate all vertices with texture

Converted Texture

Image + category

Conditional CycleGAN
Convert Image

U-Net
segmentation

Server

HoloLens

Rendering

Conversion to Camera coordinates

Projection Matrix

Texture UV coordinates

Real world

Spatial mapping meshes

HoloLens
Conclusion

◇ Summary
◇ Combined both image transformation using Deep Learning with MR on HoloLens
◇ Potential of virtual meal experience which make meal enjoyable.

◇ Future Works
◇ Implement advanced visualization and real time transformation
Application examples

- Convert diet restricted meal to your favorite meal.
- Convert simple food into luxurious meals.
System limitations

- Real time processing on HoloLens is difficult.
- A spatial map is fixed.
- A high computation cost of correspondence between each vertex and a texture taken by the camera.
Future works

✧ Implement advanced visualization.
  ✧ Improve ease of operation.

✧ Real time transformation.
  ✧ Combination of 3D reconstruction from images.
  ✧ Reduce computation cost on HoloLens side.
応用先

病院とかダイエット食とかに使えるといいかかもしれない。
QA

◇ これはなんの役にたつの？
◇ 病院食やダイエット食などの貧相な食事を視覚的にリッチな食事にすることができる。
◇ 拒食症などの場合はまともな食事を貧相にみせることも。
QA

✧ 画像のデータセットはどうしたの
  ◦ 柳井研究室ではtwitter streamを利用して食事画像を収集していた
  ◦ Alex NetをUEC FOOD 100でfine tuneしたものを利用。

✧ インタラクティブ？
  ◦ ではありません。はったら貼りっぱなし。
QA

✧ 領域分割はどうなっているの？

✧ なんでリアルタイムじゃないの？