INT3404E 20 - Image Processing: Final Report

21021491 - Ngo Thuong Hieu
21020462 - Phung Thanh Dat
21020513 - Pham Quy Duong

May 25, 2024

Abstract

This report outlines our efforts in the culmination of a Image Processing course undertaken during the second semester of our third year at the University of Engineering and Technology (UET). Our focus centered on optimizing a model for the challenging task of 2D-3D Image Retrieval. Through a series of experiments, we discovered that the SIFT algorithm yielded the most promising results among a range of image matching algorithms. Besides, we have got a score of 94.656% on the metric of MRR@5.

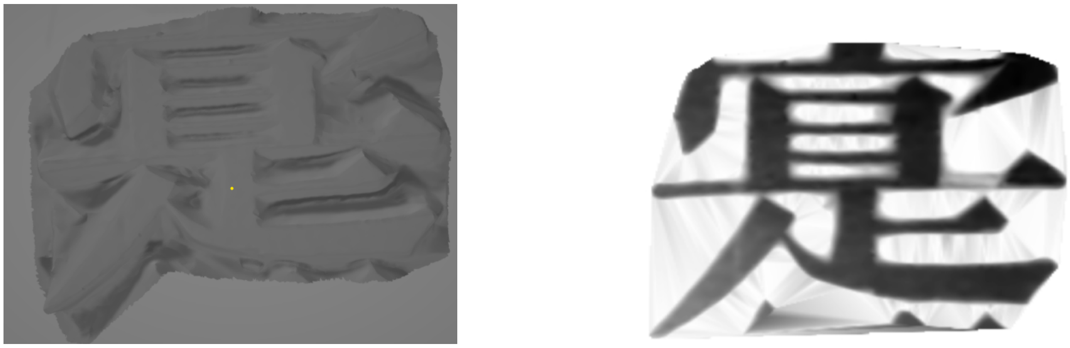
1 Introduction

This document discusses our approach to the Image Retrieval task, focusing on our method focuses on handling 3D object as well as Image matching.

The task of 2D-3D retrieval represents a significant and innovative challenge in the realm of computer vision and pattern recognition. This task involves identifying a correspondence between a two-dimensional image and its three-dimensional counterpart. Specifically, in the context of this work, the 2D image consists of a word, while the 3D object is a three-dimensional representation of that wording.

The importance of 2D-3D retrieval lies in its wide range of practical applications. From augmented reality (AR) and virtual reality (VR) environments to enhanced human-computer interaction and advanced educational tools, the ability to seamlessly match 2D textual representations with their 3D forms can revolutionize the way we interact with digital information. This capability allows for more intuitive visualization, manipulation, and understanding of data across different dimensions.

In this report, our approach involves slicing the 3D object with a plane to generate a 2D image, and then using various algorithms (e.x SIFT, ORB, BRIEF) to match the two images.



(a) Before processing

(b) After processing

Figure 1: Handling 3D object to get 2D image.

2 Proposed method

2.1 Handling 3D Object

To match a 2D image to a 3D image, we need to convert the 3D image into 2D image, which is suitable for the matching problem.

After reviewing all the 3D images, we noticed that all of which have the noisy background that is the concave ground surrounding the letter. So, we think that the noisy background need to be handled.

There is a point call "Center of Gravity" in the 3D object, which was used to draw a plane cutting through the 3D object. In our method, we defined the "Center of Gravity" as the middle of the maximum point in all three dimension: x, y, z.

We tried to extract the upper part of the plane containing wording shape as well as the surface of the object and then save it as 2D image.

2.2 Using CV2 Built-in Matching Algorithm

We have tested some matching algorithm to find the most promising algorithm to handle the task. The algorithm will count the matching points between the query image and the target image, each image in the querying database will be matched with every images in the 2D database (containing the image of 3D objects after the 3D handling phase).

Some of the matching algorithms we have adapted to our method are: SIFT, ORB, BRIEF. Figure 2 describes how the algorithm works on our task of image matching problem.

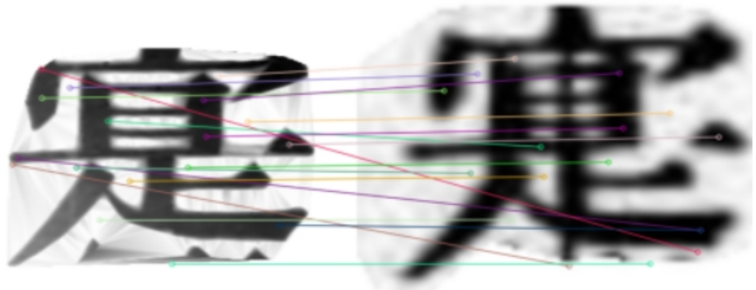


Figure 2: Image matching example.

2.2.1 SIFT

The SIFT (Scale-Invariant Feature Transform) algorithm is a computer vision technique used for feature detection and description. It detects distinctive key points or features in an image that are robust to changes in scale, rotation, and affine transformations. SIFT(scale invariant feature transform) works by identifying key points based on their local intensity extrema and computing descriptors that capture the local image information around those key points. These descriptors can then be used for tasks like image matching, object recognition, and image retrieval.

2.2.2 BRIEF

BRIEF(Binary Robust Independent Elementary Features) is an efficient and fast descriptor algorithm, ideal for real-time image matching scenarios. While it has its limitations, the trade-off between speed and robustness makes BRIEF an appealing choice for many computer vision tasks. BRIEF (Binary Robust Independent Elementary Features): This algorithm extracts descriptors from the selected keypoints. It utilizes a set of binary bits to represent the local intensity patterns around each keypoint, making it robust to noise and image distortions.

2.2.3 ORB

ORB (Oriented FAST and Rotated BRIEF) is a feature detection and descriptor algorithm commonly used in computer vision applications, particularly for image matching and object recognition. It is a robust and efficient algorithm that can handle various image transformations, such as rotation, scaling, and illumination changes. ORB is a fusion of FAST(Features from Accelerated Segment Test) keypoint detector and BRIEF descriptor with some added features to improve the performance.

2.3 Using Pre-trained CNNs

We had carried out some experiments using some pre-trained CNNs, which were ResNet50, VGG19, EfficientNet B0 to extract images feature and then computed the Cosine Similarity

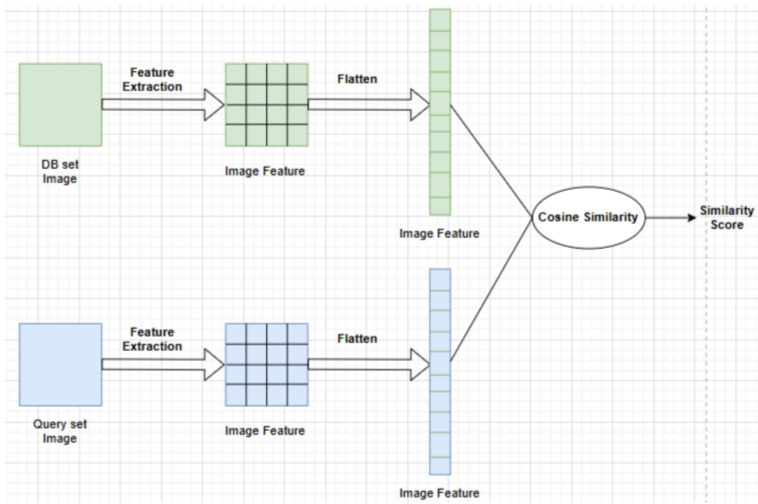


Figure 3: Image matching using pre-trained CNNs.

between an image in query set and an image in database set.

2.4 Using Image Reconstruction Model

We had built a image reconstruction model, which was an autoencoder.

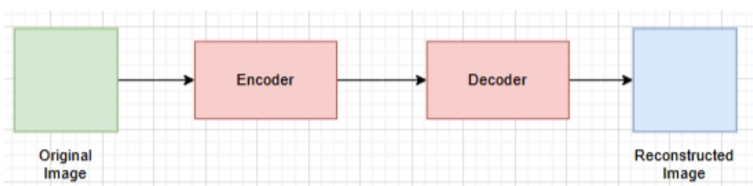


Figure 4: Image Reconstruction Model.

We then used the encoder of the model as the feature extractor for image data. Then from the extracted features, we can compute the k-nearest neighbors of the query image based on the similarity between feature vectors.

3 Experiment

3.1 Dataset

We used the dataset provided by Teacher Assistant, which contains 228 images as the querying images and 228 3D-objects as the database. More specifically, the dataset is about Sino character.

3.2 Metric

In the image retrieval problem, the Mean Reciprocal Rank(MRR) is a very common metric that helps evaluate the quality of recommendation and information retrieval systems.

MRR@5: After training the model, the input image will give some output images, and we will take top 5 images having highest score to calculate the MRR metric.

$$MRR@5 = \frac{1}{N} \sum_{n=1}^{N} \frac{1}{rank_i} \tag{1}$$

3.3 Results

Table 1. below shows the results of our methods using CV2 built-in matching algorithms. It can easily be seen that SIFT algorithm is the most promising one among the algorithms with the score of 0.947 in MRR@5.

	SIFT	**ORB**	**BRIEF**
MRR@5	0.947	0.264	0.050

Table 1: The results of our method using CV2 built-in algorithms

Table 2. below shows the results of CNNs methods including using pre-trained CNNs models as well as training a model for feature extraction. From the results, we can assure that without fine-tuning, the CNNs models cannot perform very well on new datasets

	ResNet50	**VGG19**	**EfficientNet B0**	**ImageRecons**
MRR@5	0.409	0.464	0.512	0.812

Table 2: The results of CNNs.

Besides, with just a few convolutional layers in the autoencoder, we could reach to the score of 0.812 in MRR@5.

4 Future works

Fine-tuning the CNNs models will definitely necessary to improve the performance of our method, which we think the results will be considerably increased. Moreover, the architecture of the Image Reconstruction method also has to be more complex to capture image embeddings more efficiently.

5 Conclusion

In this report, we have proposed fully detailed explanations for the task of Image Retrieval, encompassing our current methodologies and future directions. This task has provided us with substantial experience in both 3D handling and image matching techniques. Our current methods leverage advanced algorithms and machine learning models to enhance the accuracy and efficiency of image retrieval processes. We have explored various approaches, including feature extraction, similarity measurement, and deep learning techniques, to ensure precise matching and retrieval of images from extensive databases.