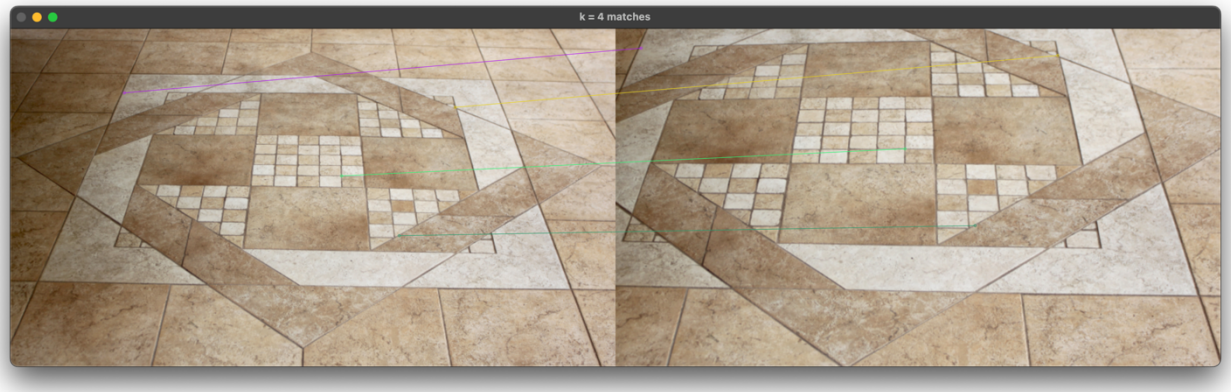
# CSIE5429 3D Computer Vision with Deep Learning Applications

## Homework 1 Report

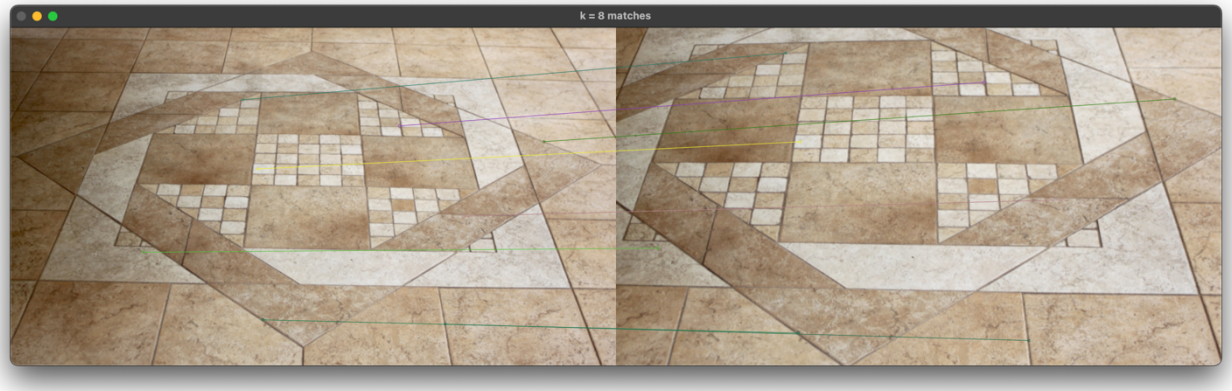
**Name:** 高榮浩 **ID:** R12922127

- **Problem 1: Homography Estimation**
  - **Image Pair #1 (Anchor A: 1-0.png & Target B: 1-1.png)**
    - ***k* = 4**



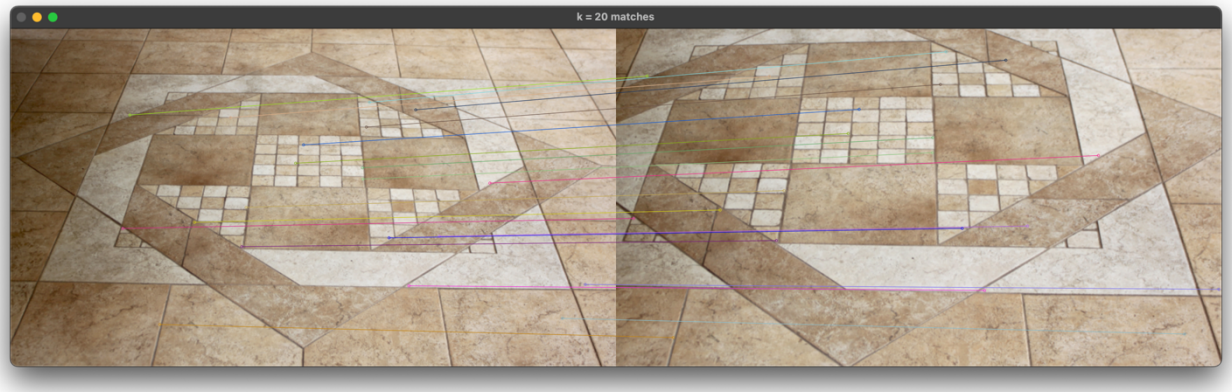

```
k = 4
Anchor sample:
[[ 989.17144775  526.9130249 ]
 [1131.79589844  200.64884949]
 [ 842.39074707  375.02770996]
 [ 289.53683472  163.97000122]]
Target sample:
[[ 914.46728516  501.24224854]
 [1125.2166748    68.93658447]
 [ 736.29412842  306.05636597]
 [  64.9059906    50.32016373]]
DLT H =
[[-3.84389220e-03  1.39477528e-04  8.81004212e-01]
 [ 1.20033265e-04 -4.08555432e-03  4.73063866e-01]
 [ 1.96787141e-07 -8.43891735e-08 -3.26432229e-03]]
DLT error = 0.05955095639755225
NDLT H =
[[ 6.99193498e-01 -2.53705817e-02 -1.60252261e+02]
 [-2.18337233e-02  7.43151177e-01 -8.60490257e+01]
 [-3.57950437e-05  1.53501603e-05  5.93771313e-01]]
NDLT error = 0.05955095669295 89
```

◆ $k = 8$




```
k = 8
Anchor sample:
[[1152.80310059  480.0057373 ]
 [ 338.53125     571.68695068]
 [1105.17980957  754.171875  ]
 [ 989.53759766  250.24342346]
 [ 628.21002197  359.78753662]
 [ 591.08911133  183.9206543 ]
 [ 641.47845459  743.29718018]
 [1356.15722656  290.51303101]]
Target sample:
[[1131.59667969  438.3850708 ]
 [ 104.92735291  561.22644043]
 [1048.74060059  796.02520752]
 [ 934.49053955  139.63697815]
 [ 467.8883667   290.78341675]
 [ 431.37039185   65.40428162]
 [ 462.62417603  776.85961914]
 [1419.8918457   182.33621216]]
DLT H =
[[-3.84163561e-03  1.38876354e-04  8.80993709e-01]
 [ 1.19424878e-04 -4.08286390e-03  4.73083486e-01]
 [ 1.96220583e-07 -8.45469195e-08 -3.26168431e-03]]
DLT error = 0.057838822717913504
NDLT H =
[[-7.08717851e-01  2.56787690e-02  1.62477694e+02]
 [ 2.20631611e-02 -7.53241884e-01  8.72442292e+01]
 [ 3.62750884e-05 -1.55207228e-05 -6.01870783e-01]]
NDLT error = 0.048659723436497625
```

◆ $k = 20$




```
k = 20
Anchor sample:
[[ 467.89974976  492.88827515]
 [1461.27905273  650.93695068]
 [1218.89099121  393.23614502]
 [ 894.98327637  354.33572388]
 [1090.18371582  528.34082031]
 [1403.14196777  737.01269531]
 [1012.93859863  654.29882812]
 [ 285.52047729  508.49865723]
 [1030.20092773  207.00320435]
 [ 599.21533203  454.85787964]
 [ 744.98339844  296.10687256]
 [ 904.80212402  250.84806824]
 [ 587.9821167   554.80529785]
 [ 379.60592651  752.59155273]
 [ 895.8951416   383.02383423]
 [ 303.39328003  220.40176392]
 [ 559.09735107  226.74441528]
 [ 912.72741699  188.02862549]
 [ 724.78314209  342.78356934]
 [ 962.41351318  532.36126709]]
Target sample:
[[ 264.28353882  461.73822021]
 [1532.44519043  663.38824463]
 [1225.99023438  323.05130005]
 [ 804.92089844  277.80673218]
 [1045.60864258  502.39663696]
 [1445.29516602  776.94763184]
 [ 936.47369385  666.08349609]
 [  44.73513031  482.97766113]
 [ 990.68157959   80.94155884]
 [ 426.63497925  412.35653687]
 [ 617.68170166  205.90982056]
 [ 824.46142578  142.59414673]
 [ 407.22174072  539.31652832]
 [ 145.34788513  785.78710938]
 [ 804.17547607  315.86682129]
 [  78.82590485  121.24275208]
 [ 389.43212891  121.98266602]
 [ 838.71722412   59.71214294]
 [ 589.7666626   267.22821045]
 [ 879.84973145  508.40264893]]
DLT H =
[[ 3.84173420e-03 -1.38797468e-04 -8.81066351e-01]
 [-1.19399967e-04  4.08244591e-03 -4.72948188e-01]
 [-1.95904442e-07  8.43010670e-08  3.26139473e-03]]
DLT error = 0.059681335774023514
NDLT H =
[[ 7.05055235e-01 -2.54597412e-02 -1.61681230e+02]
 [-2.19623039e-02  7.49327755e-01 -8.67787283e+01]
 [-3.60508608e-05  1.55524761e-05  5.98641313e-01]]
NDLT error = 0.0377227225168081
```

■ **Image Pair #2 (Anchor A: 1-0.png & Target C: 1-2.png)**

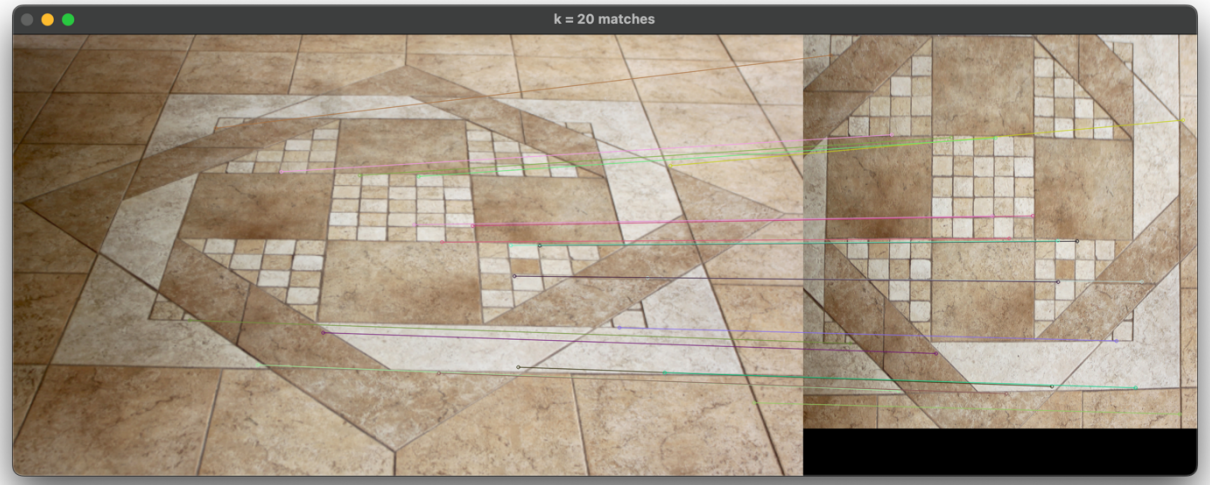
◆ $k = 4$


```
k = 4
Anchor sample:
[[1138.62695312  411.04934692]
 [1261.13720703  490.11593628]
 [ 790.43249512  691.88897705]
 [ 360.62591553  561.59747314]]
Target sample:
[[607.15252686 400.95422363]
 [671.90649414 500.54681396]
 [371.64208984 732.83331299]
 [102.49594116 606.08148193]]
DLT H =
[[-1.82188833e-03 -6.26852417e-04  7.17986246e-01]
 [ 8.28119288e-06 -4.30949773e-03  6.96038880e-01]
 [-1.02645631e-07 -1.73483093e-06 -1.82855211e-03]]
DLT error = 0.4308802849587283
NDLT H =
[[ 3.73534684e-01  1.28521115e-01 -1.47205930e+02]
 [-1.69786080e-03  8.83559572e-01 -1.42706147e+02]
 [ 2.10450348e-05  3.55685644e-04  3.74900934e-01]]
NDLT error = 0.43088028526945166
```

◆ $k = 8$


```
k = 8
Anchor sample:
[[ 463.56973267  657.32147217]
 [ 790.45013428  276.83413696]
 [ 266.17803955  552.80163574]
 [1048.98974609  235.13372803]
 [1281.1940918   255.94462585]
 [ 790.43249512  691.88897705]
 [ 477.04867554  643.71704102]
 [ 395.09732056  182.80702209]]
Target sample:
[[178.22705078 706.40484619]
 [374.19943237 202.11230469]
 [ 40.30236816 599.96362305]
 [571.92541504 132.34724426]
 [739.72479248 167.16596985]
 [371.64208984 732.83331299]
 [185.55516052 694.13861084]
 [ 53.18943024  40.92787933]]
DLT H =
[[-1.81742460e-03 -6.25879772e-04  7.17734143e-01]
 [ 4.93098733e-06 -4.29497492e-03  6.96298985e-01]
 [-1.09931798e-07 -1.73488848e-06 -1.81135367e-03]]
DLT error = 0.3648493168974574
NDLT H =
[[-4.01009039e-01 -1.38205168e-01  1.58332451e+02]
 [ 1.30008408e-03 -9.48350513e-01  1.53575498e+02]
 [-2.35905594e-05 -3.82964854e-04 -4.00472420e-01]]
NDLT error = 0.26793275001703093
```

◆ $k = 20$




```
k = 20
Anchor sample:
[[ 344.43087769  556.94787598]
 [ 476.96582031  643.63024902]
 [1236.60070801  474.22335815]
 [1444.26049805  717.37335205]
 [1026.69665527  411.4647522 ]
 [ 895.94781494  373.03851318]
 [ 977.34118652  470.94390869]
 [ 781.77819824  370.82913208]
 [1281.1940918   255.94462585]
 [ 830.2144165   659.71844482]
 [ 675.83551025  274.6781311 ]
 [ 523.04943848  267.70504761]
 [1181.60876465  571.10180664]
 [ 604.55560303  580.98083496]
 [ 836.08514404  404.74911499]
 [ 395.09732056  182.80702209]
 [ 984.7677002   648.13238525]
 [ 970.21118164  411.14651489]
 [ 790.45013428  276.83413696]
 [1269.8782959   658.6986084 ]]
Target sample:
[[ 91.40249634 601.800354  ]
 [185.55516052 694.13861084]
 [659.30584717 481.61758423]
 [733.36138916 739.50701904]
 [533.50579834 402.91375732]
 [446.80288696 353.56002808]
 [496.2204895  482.20440674]
 [368.51876831 353.91574097]
 [739.72479248 167.16596985]
 [395.45205688 699.59393311]
 [288.11401367 201.19953918]
 [172.10244751 195.64083862]
 [609.97802734 597.07983398]
 [258.56167603 621.49822998]
 [400.93927002 397.47052002]
 [ 53.18943024  40.92787933]
 [484.29406738 685.68023682]
 [495.67572021 402.36273193]
 [374.19943237 202.11230469]
 [647.1940918  688.26647949]]
DLT H =
[[ 1.81956215e-03  6.27490149e-04 -7.18238546e-01]
 [-8.32691591e-06  4.30588382e-03 -6.95778576e-01]
 [ 1.04667868e-07  1.73490593e-06  1.82186529e-03]]
DLT error = 0.2700398511500335
NDLT H =
[[-3.85378152e-01 -1.32815661e-01  1.52016013e+02]
 [ 1.92192573e-03 -9.12190813e-01  1.47243673e+02]
 [-2.19078574e-05 -3.67405758e-04 -3.86298844e-01]]
NDLT error = 0.3626585352922621
```

## ■ Error

### ◆ Image Pair #1 (Anchor A: 1-0.png & Target B: 1-1.png)

| *# of sample \ Method* | *DLT* | *Normalized DLT* |
|---|---|---|
| *k = 4* | **0.05955095639755225** | 0.0595509566929589 |
| *k = 8* | 0.057838822717913504 | **0.048659723436497625** |
| *k = 20* | 0.059681335774023514 | **0.0377227225168081** |

When *k* is set to 4, the reprojection error generated using DLT is roughly equivalent to the reprojection error obtained through normalized DLT. However, when *k* takes values of 8 and 20, the reprojection error produced by normalized DLT outperforms that of DLT. This illustrates the concept emphasized by the professor during the lecture: "The basic DLT algorithm is never used with more than 4 point-correspondences."

### ◆ Image Pair #2 (Anchor A: 1-0.png & Target C: 1-2.png)

| *# of sample \ Method* | *DLT* | *Normalized DLT* |
|---|---|---|
| *k = 4* | **0.4308802849587283** | 0.43088028526945166 |
| *k = 8* | 0.3648493168974574 | **0.26793275001703093** |
| *k = 20* | **0.2700398511500335** | 0.3626585352922621 |

The aforementioned result closely resembles the outcome observed in Image Pair #1. However, when k is increased to 20, the reprojection errors are notably higher than those when k equals 8. Additionally, it's worth noting that the reprojection errors obtained with the DLT surpass those produced by the normalized DLT. After conducting numerous trials, a consistent pattern emerged: when k equals 20, the reprojection errors are higher than when k equals 8, and DLT consistently outperforms normalized DLT in terms of reprojection error. Consequently, I believe this is an issue that warrants further investigation.

## ■ Bonus

### ◆ My Method

The following method pertains to the RANdom SAmple Consensus (RANSAC).

1. Utilize the "`get_sift_correspondences`" function to determine the correspondences.
2. Randomly select *k* pairs from the obtained matching results.
3. Estimate the homography between the anchor image and target images using both the direct linear transform and the normalized direct linear transform methods.
4. Calculate the reprojection error by comparing the matching pairs against the ground truth.
5. Calculate the number of inliers by comparing the error of each pair obtained through the use of normalized direct linear transform with a predefined threshold. Pairs with errors less than the threshold are categorized as inliers.
6. Repeat steps 2 to 5 for a total of 50,000 iterations. The time with the highest number of inliers will be the output.

## ■ Discussion

Firstly, I observe that the errors in Image Pair #1 are consistently lower than those in Image Pair #2. This discrepancy can be attributed to the fact that the number of correspondences found in Step 1 for Image Pair #2 is significantly fewer than that for Image Pair #1. Consequently, for the same value of k, the quality of the sample in Image Pair #1 is superior, leading to this outcome.

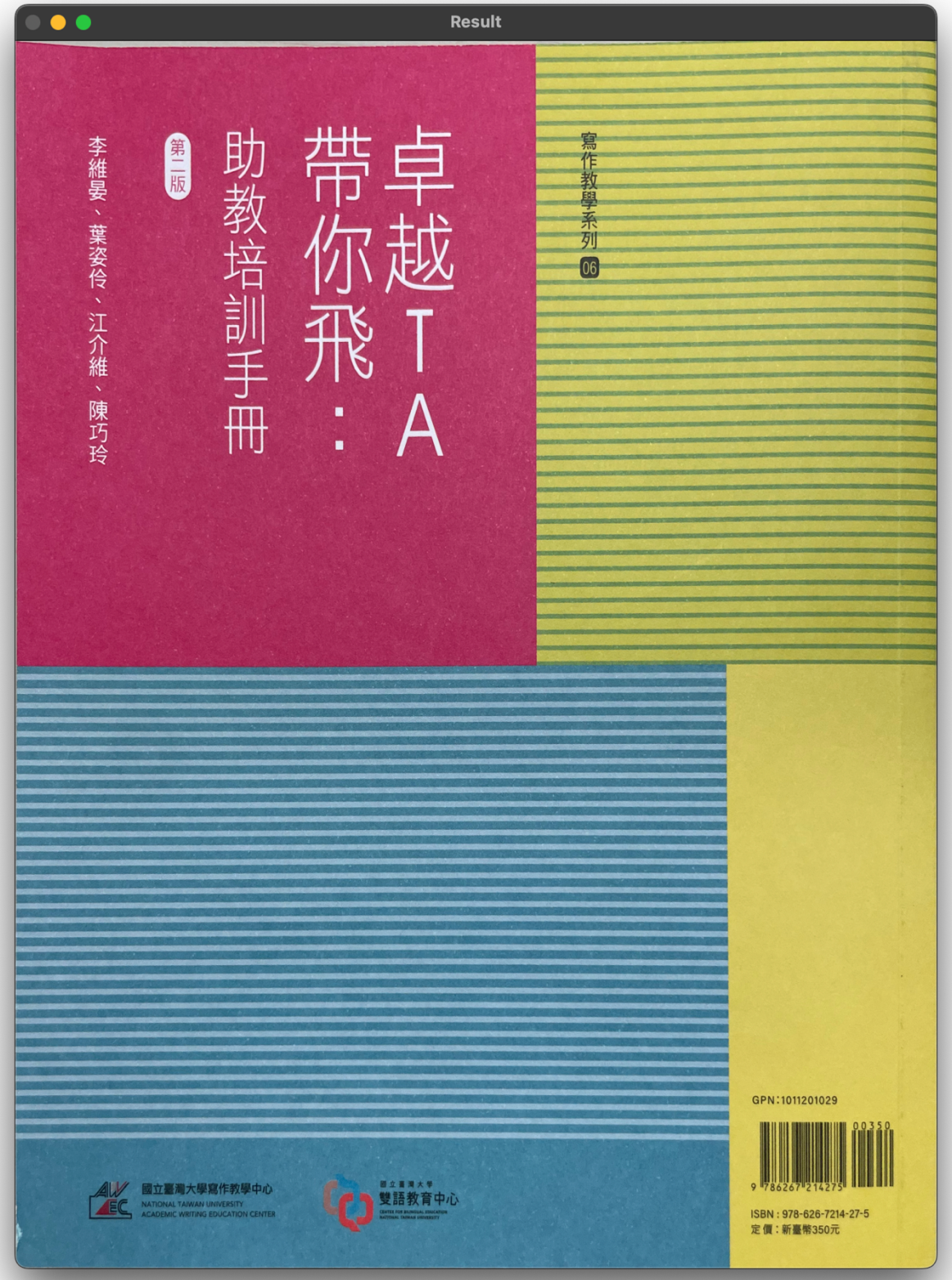
Secondly, regarding the issue mentioned on the previous page - when k is set to 20, the reprojection errors are higher than when k is set to 8, and DLT consistently outperforms normalized DLT in terms of reprojection error - I believe that this is also due to the superior quality of the sample in Image Pair #1 compared to Image Pair #2, thus resulting in this outcome.

Thirdly, the challenge I encountered pertains to determining the appropriate threshold value in RANSAC. Initially, I set the threshold value as a constant, but I found that it did not work effectively for both image pairs. Subsequently, I considered that the threshold value should vary with the size of the target image. When the target image is larger, the threshold value should be adjusted accordingly. While this approach yielded better results, it did not entirely meet my expectations. Therefore, this issue remains unresolved, and it may be necessary to explore dynamically adjusting the threshold value as a potential avenue for future work.

Fourthly, similar to the aforementioned challenge, there is also a difficulty in determining the appropriate number of iteration times in RANSAC. I empirically set it to 50,000, but it might also be a subject for future research, as mentioned previously.

- **Problem 2: Document Rectification**
  - **Input Document Image**

## ■ Rectified Result




Please note that the distortion at the top of the rectified image and the wear at the lower-left corner of the rectified image are actually inherent features present in the source image.

- **My Method**
  1. Manually mark the corner points on the image in the following sequence: Upper left, Upper right, Lower left, Lower right.
  2. Compute the homography using the normalized direct linear transform method for image transformation.
  3. Employ bilinear interpolation for image warping.

  Please note that the backward warping method is used instead of the forward warping method. Thus, when estimating the homography in step 2, I directly consider the rectified image as the anchor image and the source image as the target image.

- **YouTube Link**

  https://youtu.be/ZJcRHNkv9B0

- **Execution**
  - **Command**
    - **Problem 1: Homography Estimation**
      - **Image Pair #1 (Anchor A: 1-0.png & Target B: 1-1.png)**

        `python 1.py images/1-0.png images/1-1.png groundtruth_correspondences/correspondence_01.npy`

      - **Image Pair #2 (Anchor A: 1-0.png & Target C: 1-2.png)**

        `python 1.py images/1-0.png images/1-2.png groundtruth_correspondences/correspondence_02.npy`

    - **Problem 2: Document Rectification**

      `python 2.py images/2.png`

  - **Environment / Package**
    - `python==3.8.18`
    - `opencv==4.5.2`

      Please note that I was unable to install OpenCV version 4.5.1.48, so I opted to install version 4.5.2 as an alternative.
    - `numpy==1.19.5`