

MRZ Processing Flow Documentation

Tổng quan

MRZ (Machine Readable Zone) là vùng dữ liệu máy đọc được ở mặt sau của CCCD mới. API này xử lý và trả về thông tin MRZ cho frontend.

Flow xử lý MRZ

1. Upload ảnh

```
POST /api/scan/  
Body: multipart/form-data với file ảnh
```

2. Detection card type

```
# Detect loại thẻ bằng YOLO  
detected_label in:  
- 'cccd_qr_front'      # CCCD có QR - mặt trước  
- 'cccd_qr_back'      # CCCD có QR - mặt sau  
- 'cccd_new_front'    # CCCD mới - mặt trước  
- 'cccd_new_back'     # CCCD mới - mặt sau (CÓ MRZ)
```

3. Process MRZ (chỉ cho back side)

Case 1: CCCD có QR (cccd_qr_back)

```
if detected_label in ['cccd_qr_front', 'cccd_qr_back']:  
    ocr_processor = OCR_CCCD_QR(face=detected_label)  
    ocr_result = ocr_processor.process_image(temp_path)  
  
# Process MRZ chỉ khi là back side  
if 'back' in detected_label:  
    mrz_result = ocr_processor.process_mrz(temp_path)
```

Case 2: CCCD mới (cccd_new_back)

```
elif detected_label in ['cccd_new_front', 'cccd_new_back']:  
    ocr_processor = OCR_CCCD_2025_NEW()  
    ocr_result = ocr_processor.process_image(temp_path)
```

```
# Process MRZ chỉ khi là back side
if 'back' in detected_label:
    mrz_result = ocr_processor.process_mrz(temp_path)
```

Response Structure

Success Response (có MRZ)

```
{
  "status": "completed",
  "message": "Image processed successfully",
  "task_id": "uuid-here",
  "timing": {
    "start_time": 1759506800.9486952,
    "end_time": 1759506803.144318,
    "start_timestamp": "2025-10-03T15:53:20.948697",
    "end_timestamp": "2025-10-03T15:53:23.144349",
    "total_elapsed_time": 2.196
  },
  "image_info": {
    "original_size": 706873,
    "load_method": "PIL",
    "format": "JPEG",
    "quality_score": 96.53,
    "width": 2048,
    "height": 1352,
    "blur_score": 472.53,
    "brightness": 151.35,
    "contrast": 46.82
  },
  "results": {
    "processing_time_sec": 2.196,
    "results": [
      {
        "id": "",
        "name": "",
        "birth": "",
        "sex": "",
        "nationality": "",
        "place_of_origin": "",
        "place_of_residence": "",
        "expiry": ""
      }
    ]
  },
  "details": [
    {
      "card_info": {
```

```

    "detections": [...],
    "image_quality": {...},
    "debug_info": {...}
  },
  "card_ocr_results": {
    "full_name": "",
    "id_number": "",
    "date_of_birth": "",
    "sex": "",
    "nationality": "",
    "place_of_origin": "",
    "place_of_residence": "",
    "expiry": "",
    "date_of_issue": ""
  },
  "timing": {},
  "ocr_fields_count": 1,
  "ocr_fields_total": 13
}
],
"mrz_result": {
  "status": "success",
  "message": "Found 4 MRZ text lines",
  "texts": [
    "G>>0022002072092002009600207",
    "9901203M3901201VNM<<<<<<<<<<4",
    "S<",
    "LF<<CONG<ANH<<<<<<<"
  ],
  "mrz_string":
"G>>00220020720920020096002079901203M3901201VNM<<<<<<<<<<4S<LF<<CONG<ANH<<<<<<<
<",
  "mrz_length": 79,
  "total_mrz_regions": 1,
  "dates_found": [],
  "total_dates": 0,
  "all_ocr_texts": [
    "G>>0022002072092002009600207",
    "9901203M3901201VNM<<<<<<<<<<4",
    "S<",
    "LF<<CONG<ANH<<<<<<<"
  ]
},
"start_time": 1759506800.9486952,
"elapsed_time": 2.196
}

```

Success Response (không có MRZ - front side)

```
{
  ...
  "mrz_result": {
    "status": "no_mrz_detected",
    "message": "No MRZ regions detected in the image.",
    "texts": [],
    "mrz_string": "",
    "mrz_length": 0,
    "total_mrz_regions": 0,
    "dates_found": [],
    "total_dates": 0
  }
}
```

MRZ Result Fields Explanation

Field	Type	Description
<code>status</code>	string	"success" hoặc "no_mrz_detected"
<code>message</code>	string	Mô tả kết quả
<code>texts</code>	array	Mảng các dòng text MRZ được detect
<code>mrz_string</code>	string	Chuỗi MRZ đầy đủ (concatenated)
<code>mrz_length</code>	int	Độ dài chuỗi MRZ
<code>total_mrz_regions</code>	int	Số vùng MRZ được tìm thấy
<code>dates_found</code>	array	Các ngày tháng được extract từ MRZ
<code>total_dates</code>	int	Số lượng dates tìm thấy
<code>all_ocr_texts</code>	array	Tất cả text OCR từ vùng MRZ

Testing

Test với Postman/cURL

Request

```
curl -X POST "http://localhost:8000/api/scan/" \
  -H "accept: application/json" \
  -H "Content-Type: multipart/form-data" \
  -F "file=@cccd_new_back.jpg"
```

Expected Log Output

Frontend Integration

Troubleshooting

Nguyên nhân:

- PROFESSEUR : M.DA ROS

Giải pháp:

1. Kiểm tra `detected_label` phải chứa "back"
2. Verify OCR processor implementation
3. Kiểm tra chất lượng ảnh (`blur_score`)

Issue 2: MRZ result là None trong response

Nguyên nhân:

- Biến `mrz_result` bị gán lại `None` sau khi xử lý
- Logic điều kiện sai (kiểm tra wrong label)

Giải pháp:

☑ Code đã được fix:

```
# ĐÚNG: Khởi tạo mrz_result = None ở đầu
mrz_result = None

# Xử lý MRZ trong các nhánh if/elif
if 'back' in detected_label:
    mrz_result = ocr_processor.process_mrz(temp_path)

# KHÔNG gán lại mrz_result = None sau khi xử lý
# ✗ mrz_result = None # Dòng này đã bị xóa

# Trả về response với mrz_result
"mrz_result": mrz_result if mrz_result else {
    "status": "no_mrz_detected",
    ...
}
```

Issue 3: Backend log có MRZ nhưng frontend không nhận được

Nguyên nhân:

- Response structure không đúng
- Frontend parsing sai field name

Giải pháp:

1. Check response structure: `result.mrz_result`
2. Verify JSON serialization
3. Check network tab trong browser DevTools

Code Changes Summary

Before (Bug)

```
# Line 199: Khởi tạo mrz_result trong response
mrz_result = None

# Line 201-203: Xử lý MRZ
if 'ccd_new_back' in detected_label: # ✗ SAI: kiểm tra ccd_new_back cho
ccd_qr
    mrz_result = ocr_processor.process_mrz(temp_path)

# Line 210-212: Xử lý MRZ cho ccd_new
mrz_result = ocr_processor.process_mrz(temp_path) # ✗ Xử lý cả front lẫn
back

# Line 247: GÁN LẠI NONE (BUG CHÍNH)
mrz_result = None # ✗ Làm mất dữ liệu MRZ đã xử lý
```

After (Fixed)

```
# Line 161: Khởi tạo mrz_result = None sớm
mrz_result = None

# Line 200-202: Xử lý MRZ cho ccd_qr_back
if 'back' in detected_label: # ✓ ĐÚNG: kiểm tra 'back' trong label
    mrz_result = ocr_processor.process_mrz(temp_path)

# Line 212-214: Xử lý MRZ cho ccd_new_back
if 'back' in detected_label: # ✓ ĐÚNG: chỉ xử lý back side
    mrz_result = ocr_processor.process_mrz(temp_path)

# ✓ Không gán lại mrz_result = None
# mrz_result giữ nguyên giá trị đã xử lý
```

✓ Verification Checklist

- ✓ MRZ được xử lý chỉ cho back side
- ✓ `mrz_result` không bị gán lại `None`
- ✓ Response structure có field `mrz_result`
- ✓ MRZ result có đầy đủ fields: status, message, texts, mrz_string, etc.
- ✓ No syntax errors
- ✓ Log output hiển thị MRZ result
- ✓ Frontend có thể access `response.mrz_result`

🔗 Next Steps

1. **Test với nhiều ảnh CCCD back side** để verify accuracy
2. **Parse MRZ data** để extract thông tin:

- Document number
- Date of birth
- Date of expiry
- Nationality
- Full name

3. **Validate MRZ checksum** để đảm bảo tính toàn vẹn dữ liệu

4. **Error handling** cho các trường hợp MRZ bị lỗi format

Last updated: October 4, 2025