Accredited by National Journal Accreditation (ARJUNA) Managed by
Ministry of Research, Technology, and Higher Education, Republic Indonesia with Second Grade (Peringkat 2, Sinta 2)
since year 2017 to 2021 according to the decree No. 10/E/KPT/2019.

Published online on the journal's webpage: **http://jurnal.iaii.or.id**

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 5 No. 4 (2021) 783 - 793 ISSN Media Elektronik: 2580-0760

Implementation of Verification and Matching E-KTP with Faster R-CNN and ORB

Muhammad Muttabi Hudaya[1], Siti Saadah[2], Hendy Irawan[3]
[1,2,3]Informatics, Telkom University
[1]mmhudaya@student.telkomuniversity.ac.id, [2]sitisaadah@telkomuniversity.ac.id, [3]hendyirawan@telkomuniversity.ac.id

Abstract

An EKTP image repository can be a helpful tool to assist human operators in EKTP image pair checking. But, such a repository needs a solid validation that has verification and matching uploaded images. To solve this problem, this paper implementing a detection model using Faster R-CNN and a matching method using ORB (Oriented FAST and Rotated BRIEF) and KNN-BFM (K-Nearest Neighbor Brute Force Matcher). The goal of the implementations is to reach both an 80% mark of accuracy and prove matching using ORB only can be a replaced OCR technique. The implementation accuracy results in the detection model reach mAP (Mean Average Precision) of 94%. But, the matching process only achieves an accuracy of 43,46%. The matching process using only image feature matching underperforms the previous OCR technique but improves processing time from 4510ms to 60m). Image matching accuracy has proven to increase by using a high-quality dan high quantity dataset, extracting features on the important area of EKTP card images.

Keywords: Detection, Matching, Identity Card, EKTP, ORB, Faster R-CNN.

1. Introduction

In Indonesia, using a national identity card called EKTP (Kartu Tanda Penduduk Elektronik) is very substantial. Various business processes require using EKTP, such as registering the marriage, buying a house, applying for a job, and even using medical insurance [1][2][3]. In 2013, 64.5% of that business process started to accept EKTP recorded in Palembang [4]. Using EKTP usually gives EKTP directly to the service operator or provides the EKTP with a photocopied document to a substance.

Given EKTP document then will be checked usually by a person service operator. A person operator not occasionally flawlessly constantly performed and can sometimes make a human-error mistake. Even when inserting the field data on EKTP, an error is still recorded in Gorontalo [5]. In addition, in Cibeuying Kaler recorded from 100 citizen respondents to an EKTP survey satisfaction, 84,9% of the respondent is not satisfied with EKTP because there is still a human-error mistake when inputting the data fields. [6].

There is a chance a person operator will make a human-error mistake in the future. This problem can be solved by implementing a machine-assisted for checking the EKTP document in the field of computer vision. The first process will be giving the EKTP images to a specific place that the machine could check. In the research conducted by Kevin Akbar, this problem can be solved by building a repository for an ID card to accommodate a repository for further checking by a computer [7]. But the repository needs a solid validation for whether the uploaded EKTP image is correct to minimize any frauds that can happen.

To mitigate fraud, a pair of images will be uploaded to the repository in a predefined format. The first is a selfie image of a person holding an EKTP card of theirs. And the second will be the image of the EKTP card itself. Images files that will be uploaded need to be valid, and therefore the verification of these two images needs to be implemented. The verification that will be implemented is to check whether the EKP is on the images or not. Computer vision's object detection field can solve this verification implementation.

Regarding the classification of identity documents in the research conducted by Pere Vilás, the classification model was built using the CNN architecture and succeeded in obtaining a classification accuracy of 98% [8]. In this study, the classification will not be carried out but will be built as a model for EKTP detection. The

Accepted by the Editor: 01-06-2021 | Completed Revision: 21-08-2021 | Published Online: 25-08-2021

detection method will also be built with the CNN architecture, namely Faster R-CNN. Faster R-CNN has a faster speed than its predecessor R-CNN and Fast R-CNN, with the same mean average precision (mAP) of 66.9% for the 2007 VOC dataset [9].

The validation of the images did not stop from detecting whether the EKTP is on the images. After that, we need to check whether both EKTP cards are the same from both image pairs. We can use the Faster R-CNN model to segment the EKTP from the image pair. We can solve the matching process by using OCR (Optical Character Recognition). In research conducted by Firhan Maulana, extracting the data fields inside the EKTP card reached an F-Score of 0.78 with the timely processing of 4510 milliseconds per ID card [10]. Furthermore, Pratama et al [11], improving the OCR technique with CNN and increased the average F-Score of 0.84. However, various images with good and bad quality become a challenge for the OCR engine in EKTP card data extraction[10][11]. In research conducted by Tom Yeh and Boritz Katz, combining OCR and merging image features in the study case of finding image documents has boosted its retrieval precision [12]. The study case in Tom Yeh and Boritz Katz's research has the level of complexity higher than ours, but the method of using only image matching on extracted image features looks promising in the case of EKTP matching process implementation. But it is still unclear which image feature extraction algorithms are the best in our study case.

Besides, the research conducted from [13] and [14][13] compares the different methods of image feature extraction such as SIFT, ORB, BRIEF, and SURF. ORB has proven to be the most efficient matching method after the features were extracted and overall extracting much more features from the images. Furthermore, in this study, the matching process will be implemented using the ORB algorithm for extracting image keypoints and image features. And the matching process will be completed by matching the K-Nearest Neighbor Brute Force Matching (KNN-BFM) methods [15] by finding the closest distance in 2 images keypoints.

Based on previous studies, thus, in this research, the implementation of verifying and matching will be carried out by Faster R-CNN, ORB, and KNN-BFM [9]-[15]. The validation will be consisted of verifying and matching EKTP on the image pair. Verifying validation by EKTP detection and image matching process will be repository validation for future uploaded images. The detection process in this research will carry out verification of EKTP. The detection process can be implemented using a model from the Faster-RCNN method for object detection from EKTP image pair data. In comparison, the Faster R-CNN method has proven to be one of the fastest and accurate than its predecessor [9].

ORB and KNN-BFM will carry out the EKTP image matching process. The method chosen, proven to be quite promising for implementing matching EKTP images segmented from the verification (detection) process without using any OCR methods. The aim of the study is to implement a verification model to detect whether EKTP on image pair and implementing image matching from the segmented image that gotten from the Faster R-CNN EKTP detection model using ORB and KNN-BFM and to observe does feature matching method without OCR is sufficient to carry out the matching task is one of the goals of the novelty of this study. The implementation result will need to be evaluated through its accuracy to determine whether the model and match process is achieving an accepted accuracy value above 80% to be implemented in a real-world environment.

Research Limitation

Due to the unfortunate timing of the study that happened on the global pandemic (COVID-19), the dataset of the EKTP image pair will be collected in an uncontrollable environment. In this study, the image dataset is collected through various camera-phone that resulted from a wide range of various image resolutions. The image will be submitted through an online form that the writer prepared, and the respondent will upload the pair of images through the form.

The hardware used to complete the implementation process, such as model training and image processing, will be Nvidia RTX 2070 GPU, 16GB DDR4 RAM, and Intel i5-8400 CPU.

2. Research Method

These study purposes are to successfully implementing detection and matching to helping verification the image pair uploaded. Faster R-CNN will complete implementation of the verification model. The model will help the matching process by segmenting the EKTP image for further feature extraction using ORB and matching the features using KNN-BFM shown in Figure 3. The result of implementation will be evaluated based on the total feature matched in the data image pair.

2.1. E-KTP data collection

Indonesia citizen identity card or E-KTP is compulsory to have for citizens above 17 years old. E-KTP consists of specific fields that determine the information of the owner of the card, such as citizen number, name, address gender, owner photo, handwritten owner signature, etc. The shape of E-KTP shown in Figure 2 is a rectangular card with the owner's confidential identity document.

DOI: https://doi.org/10.29207/resti.v5i4.3175

Research on an identity document problem will have issues on how the researcher will get the data. Identity documents contain sensitive personal information, which will challenge the researcher to collect data from an uncontrollable environment, as in the research by Arlazarov mentioned [16]. The research conducted by Arlazarov was collecting a dataset consist of labeled video and images (such as text segmentation, optical character recognition, forensics, etc.) [16].. Video data on this study will be more impactful on training the model, but collecting it face to face will be difficult. Hence, the data will be only collected as images from an online form and label manually.

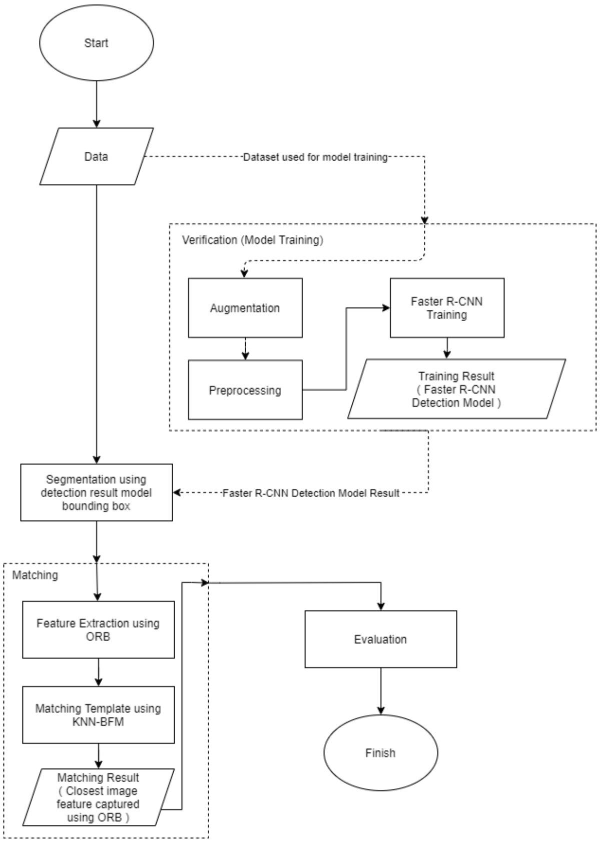

Figure 1. Flowchart of the research process

Figure 2. E-KTP

2.2. Building the datasets

The dataset was collected through an online form and asked the respondent to upload a pair of images as shown in Figure 2 and Figure 3. In Figure 3, the respondent must hold the identity document card and show their faces (selfies). And in Figure 2, the respondent needs to upload the image of the identity document card that they were holding on the previous selfie image.

Figure 3. Selfie image with identity document card

The dataset consisted of a total of 125 pairs of images. After the process, the data is labeled using COCO format and drawn in the image's bounding box. After that, remove the EXIF on the image first to ensure the metadata does not rotate.

The data will be augmented to increase the dataset count before the dataset goes into the dataset library for training machine learning models. In general, the machine learning process is directly proportional to many existing datasets, which means that the more datasets, the "better" the machine learning model will be [17]. However, collecting countless identity document card images in the real world is not easy, so data augmentation is needed to help improve the dataset to support the learning process.

Data augmentation can be a way out of performance and overfitting problems in a machine learning model. Many data augmentation techniques can be applied to a dataset, such as flip, rotation, crop, scale, or whitening [18].

This process is proven to be good enough to improve performance in cases on various datasets [16]. There is a selfie image holding an identity card in this study, where there is a possibility of differences in holding an identity card. Therefore, it is necessary to augment data with perspective transformation techniques [19]. This technique will perform an image transformation, as shown in Figure 4.

DOI: https://doi.org/10.29207/resti.v5i4.3175

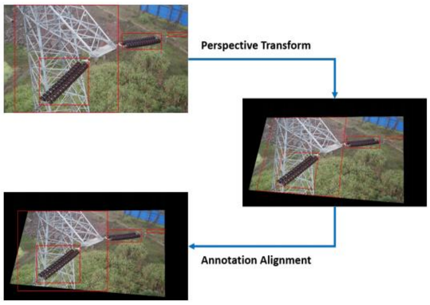

Figure 4. An illustration for perspective transformation

2.3. Faster R-CNN Detection

Faster R-CNN is a development of Fast R-CNN; the main difference with Fast R-CNN is that Fast R-CNN uses selective search to make area proposals while Faster R-CNN uses Region Proposal Network (RPN). RPN is used for making regional and network proposals that generate area proposals to handle objects detection later. This results in a shorter time for regional proposals in the Regional Proposal Network than selective tracing. The Region Proposal Network works by sorting the area boxes (anchors) and doing which ones are likely to be covered by the object. The Faster R-CNN consists of 2 modules; the first is a deep convolutional network used to make regional proposals or what is commonly called the Region Proposal Network, and the second is Fast R-CNN. Still, it only functions as a detector [9].

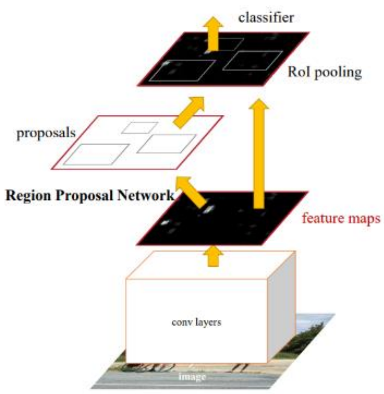

Figure 5. An illustration of Faster R-CNN Architecture

Machine learning model accuracy is directly proportional by the quantity of the dataset. In this study, we will only be using a 125 pair image dataset trained by the Faster-RCNN method. The Faster R-CNN method proves to be good for training on a small dataset [20]. The machine learning model will be trained using phyton programing language and detectron2 framework in the identity document card detection process. Detectron2 [20] is a module from Facebook with the weight of pre-trained Faster R-CNN architecture with the same base model as the original paper proposed [9].

In this study, the model will be build using a pre-trained model from ResNet-50 C4 Architecture.

2.4. Oriented Fast and Rotated BRIEF (ORB) Feature Extraction

ORB [22] is a method that is a development of the previous method, namely FAST and BRIEF, so that it has advantages in detection speed and resistance to rotation and noise [23]. The FAST technique is used at an early stage to determine the Keypoint. FAST does not calculate the orientation and rotation of the variants but calculates the intensity of the centroid patch. In ORB, matrix rotation is calculated using the orientation of the patch, and then the BRIEF descriptor guides an orientation.

Suppose the keypoint is called *Kp*. *Kp* points are used in the FAST technique to get other *Kp* points. In the initial stage, to determine the points of *Kp*, the FAST technique was used. However, FAST cannot determine the orientation and rotation of the *Kp*. FAST only calculates the centroid intensity value of the patch. The direction of the vector at the vertex towards the centroid will give the orientation of the point *Kp*. Moments are calculated to increase the rotation invariance.

In the FAST technique, the following formula (1) is used to help find the corner points in the image.

$$C = \left(\frac{m_{10}}{m_{00}}, \frac{m_{01}}{m_{00}}\right) \tag{1}$$

Then the moment value is calculated by the following formula (2). And therefore, this formula (2) ensures that the moment value calculates the x and y values with a different radius from the center point to the corner point found on (1).

$$m_{pq} = \sum x^p y^q I(x, y) \tag{2}$$

To rotate the main axis at various angles, the technique that will be used is BRIEF. In the BRIEF technique, the final feature result is a vector of 256, which can be computed with various intensity tests. Each descriptor is obtained from making binary comparisons of 2 randomly selected pixel points. This process can be calculated using the following formula (3)

$$T(p; x, y) := \begin{cases} 1 : p(x) < p(y) \\ 0 : p(x) \geq p(y) \end{cases} \tag{3}$$

Where p (x) and p (y) on (3) are the intensity at a point pixel x. With different pairs of x and y points, the descriptor, in BRIEF, will be made as a text string of n bits as follows on the following formula (4)

$$f_n(p) = \sum_{1 \leq i \leq n} 2^{i-1}\, T(p; x_i, y_i) \tag{4}$$

DOI: https://doi.org/10.29207/resti.v5i4.3175

This ORB technique is used as feature extraction from the segmented image to match the identity document card image input. ORB technique is proven to have good performance in various test cases [13], which is expected to solve the problem domain of this study.

2.5. K-Nearest Neighbor Brute Force Feature Matching

K-Nearest Neighbor Brute Force Matching (KNN-BFM) technique will carry out the matching process in this study. The KNN-BFM process can be described in Figure 6.

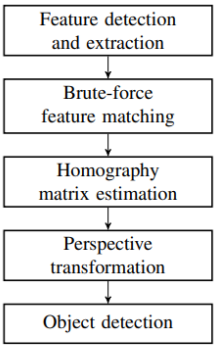

Figure 6. An illustration of the process in Brute Force Matching

The feature extraction stage in an image can be done by various methods such as SIFT, SURF, and ORB, which can then be carried out by the KNN-BFM process [15]. The process of calculating the KNN distance can use Euclidean or Hamming.

The following process that is done is to do object detection using the Homography technique. Where to look for points with the same image features but with different perspectives, the perspective transformation process is carried out in the image [19].

In this study, the pair will be brute-forced to match each data collection and the feature on each image to determine whether the images have the same features. We will separate the total matches and total "good" matches using the distance of 64 on the KNN-BFM distance result.

2.6. Evaluation

There will be an evaluation of each detection and matching process in this study.

Mean average precision (*mAP*) [24] is used as the main parameter in the Faster R-CNN modeling results. Average The precision, in this case, summarizes the comparison of recall and precision curves. A recall is a value obtained by comparing the positive value sample with the existing ground-truth value. Precision is the comparison of the positive value in the sample with the predicted results.

Average Precision calculates the maximum precision value for each recall value on the N available data in the formula 5.

$$\text{Average Precision} = \frac{1}{\mathrm{N}}\sum_{\mathrm{r}} \mathrm{APr} \quad (5)$$

Where r is the recall value for each data, the results are then interpolated by taking the maximum precision value that exceeds the value of r.

The results of average precision from formula (5) are then compared with the bounding box of the confidence in the system. The detection results are determined by ground truth and are assessed based on true and false positive values calculated by overlapping or overlapping boxes.

$$ao = \frac{area(Bp \cap Bgt)}{area(Bp \cup Bgt)} \quad (6)$$

To determine that the detection results are correct, the overlap ratio (*ao*) between the bounding box predicted by Bp and the bounding box ground truth (*Bgt)* must be above 50%, as shown in the formula (6)

A matching step will follow the results of the identity card detection on an image. This step means finding whether each pair dataset matches with each other or even matches with other images that are not the same pair. In this study, using accuracy from the confusion matrix that is shown on this formula is enough

$$Accuracy = \frac{TP + TN}{TP + TN + FT + FN} \quad (7)$$

Each dataset pair will be tested to find whether the dataset is matched with other different datasets. After extracting the features of the images, steps followed by matching algorithm and deciding the accuracy from formula (7) using a threshold that determined later.

3. Result and Discussion

In this section of the study, there are two results of EKTP verification and implementation of EKTP matching. The implementation of verification will be completed by training a Faster R-CNN model to detect EKTP on dataset images. The implementation of matching will be completed by image feature matching using the ORB algorithm and KNN-BFM. The implementation results will be evaluated with mAP on Faster R-CNN verification model and accuracy from confusion matrix on EKTP matching implementation. Implementation result on EKTP matching also will be compared to previous studies with OCR methods to see a better grasp of the analysis. As the research method has been done, the methods chosen are expected to produce promising results.

DOI: https://doi.org/10.29207/resti.v5i4.3175

3.1. Implementation of EKTP Image Pair Verification

Before the dataset feed into the Faster R-CNN training model, there is no preprocessing on the data, but there is a data augmentation process since there are only 125 datasets. The data augmentation process consists of a 75% chance of random flip, a 25% chance of random crop, a 50% chance of random rotation, and adding 25% of random brightness with 0.9 – 1.1 value.

The model was trained using detectron2 and with a pre-trained model from ResNet-50 C4 architecture. After tweaking the configuration to detect identity document cards such as 0.00025 learning rate, 2 images per batch, 100 batch size per image, 1250 iterations, and 1 num classes.

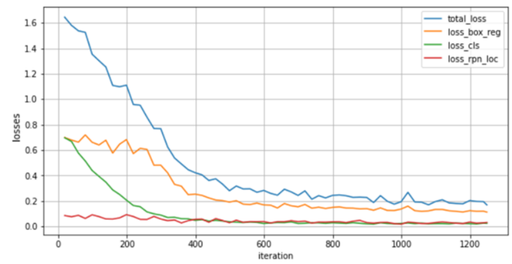

Figure 7. Graph of model losses on 1250 iterations

From Figure 7. Graph of model losses on 1250 iterations, the model reached an mAP of 91%. The model configuration was updated to 3000 iterations and reached 94%, as Figure 8. Graph of model losses on 3000 iterations stated. After several tweaks on model configuration such as learning rate, batch size, and images per batch, the model seems cannot achieve more than 94%.

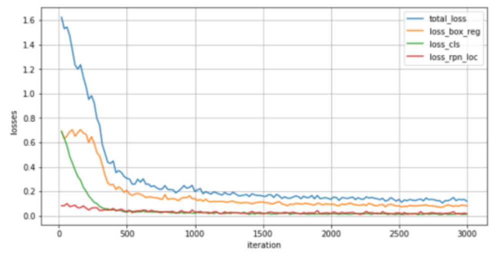

Figure 8. Graph of model losses on 3000 iterations

After the output model is generated, the model will segment the dataset into chunks of segmented document identity cards by segmenting the image from the resulting bounding boxes shown in Figure 9.

Figure 9. Example of segmented image result.

Figure 10. Example of image with both false and true positive results

Figure 10 shows the example of some false positive segmented image that is not a document identity card, but the images can be verified through the image matching process.

Figure 10. Example of image with both false and true positive results

3.2. Implementation of EKTP Image Pair Matching

In this process, the implementation of matching will be carried out by ORB and KNN-BFM. Figure 11 shows the result of matches from both methods in the plot of drawing a line on closest match features on two segmented images. The closest match features will be counted on each image.

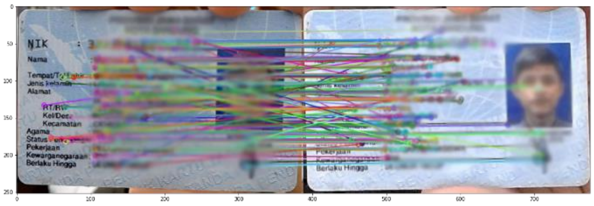

Figure 11. A plot of 2 images feature matching

The implementation result will be evaluated with two scenarios that experimented to achieve accurate evaluation. Scenario 1 (S1) consists of matching the pair of the segmented dataset images using a matching algorithm. And Scenario 2 (S2) consists of matching each segmented dataset image to any other dataset image except its pair. All result in S1 is stated as true positive and result in S2 is stated as true negative.

Table 1. 5 Sample result of matching process

Query Image Id	Train Image Id	Total Good Matches
7_card_0	7_selfie_0	515
100_card_0	100_selfie_0	317
38_card_0	38_selfie_0	764
78_card_0	78_selfie_0	636
78_card_0	78_selfie_1	3

As shown in Table 1, the distribution of total good matches of each scenario will be plotted. Finding the right threshold of matches is the target of this process of plotting. Before producing the distribution plot, removing outliers is essential to achieve a good quality of the plot.

DOI: https://doi.org/10.29207/resti.v5i4.3175

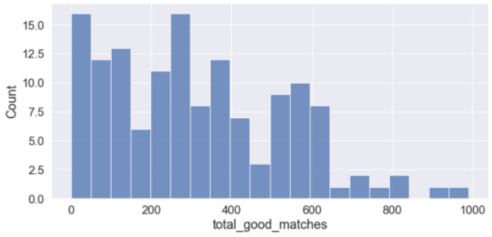

Figure 12. Distribution total matches of S1 (true positives) images using ORB

As shown in Figure 12 and Figure 13, the distribution of total matches is not centralized on a particular point that can determine the image as match or not to each other. Because the distribution is not centralized, determining the threshold by taking the median of mean in S1 and mean in S2 as the threshold. The accuracy of the matching process is by the threshold of 346 matches achieve 43.46%.

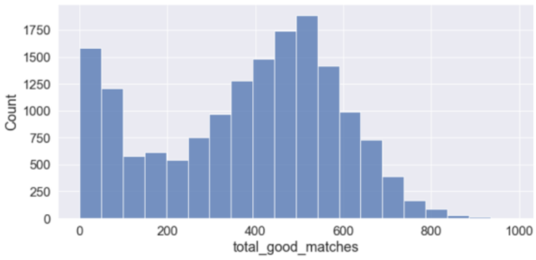

Figure 13. Distribution total matches of S2 (true negatives) images using ORB

3.3. OCR and ORB Image Matching Comparison

As Table 2 shown, the process matching using just ORB matching algorithm does not achieve the same result as previous studies using the OCR technique. In previous studies by Firhan Maulana, the OCR technique achieve an F-Score of 0.78 on overall camera conditions, but the processing time takes an overall of 4510 milliseconds per card [9]. F-Score of using only ORB image matching does not come close to OCR technique F-Score result. But if we can improve the F-Score and get a better threshold of algorithm matches, room for improvement in processing time will be the next step. The implementation of matching using ORB algorithms takes 60ms of processing time. This result of matching implementation significantly improves in terms of processing time from the OCR technique by 98.6%.

Table 2. OCR and ORB comparison

	F-Score	Processing time (per card)
OCR Technique	**0.78**	4510ms
ORB Image Matching	0.21	**60ms**

Discussions

The study's experimental results show that they are still not achieving their full potential because of some factors. Since we need to achieve an 80% mark on the accuracy, the matching process is still far from the minimal mark to be accepted on the real-world environment implementation. The implementation of Faster-RCNN detection model mAP achieves 94%, and the matching process using ORB achieves 43.46%. The comparison between matching using OCR technique and with ORB image matching only also resulted with under expectation result of F-Score.

This section of the study will analyze why the accuracy is below the mark expectation in both implementations and determine how to improve the image matching process for replacing the OCR technique. The analysis can be a good reference for future studies to improve our chosen method as a newly proposed method.

Quantity and Quality of Dataset

On the matching process implementation, image matching results achieve 43.46% accuracy using the ORB algorithm that is below the expectation of the study. From S1, we will try to sort ascendingly to analyze the lowest total matches in the image dataset and observe the image matched quality.

We determine the lowest total match by manually selecting the false EKTP detection object threshold with the highest feature matched, as shown in Figure 14. After that, we filter out the dataset with the threshold to analyze the true-positives dataset that has low image feature total good matches.

Figure 14. False EKTP segmented images from verification (Faster R-CNN) machine learning model

Table 3. 5 lowest feature matched on S1sceneario

Query Image Id	Train Image Id	Total Good Matches
120_card_0	120_selfie_0	1
18_card_0	100_selfie_0	1
53_card_0	53_selfie_0	19
46_card_0	46_selfie_0	20
122_card_0	122_selfie_0	30

As Table 5 shown, there is still a true-positives image with low total good matches. There is a total of 11 image datasets filtered from the threshold of false EKTP objects. Usually, losing 11 pairs of datasets is tolerable, but 11 images are equal to almost 10% of the collected dataset in our study.

DOI: https://doi.org/10.29207/resti.v5i4.3175

Figure 15. Example of blurry image from dataset

Figure 16. Example of low-resolution segmented image from dataset

We subjectively determine its causes by discovering EKTP content in the lowest dataset feature matched dataset measured if the content is humanly visible or not. Figure 15 and Figure 16 are examples of determining if the EKTP image content is not humanly visible. One reason the image dataset with low total good matches from S1 was blurry images, camera noises, and low-resolution segmented EKTP.

Image Feature Extraction Algorithm

We tried to use another feature extraction algorithm such as SIFT to determine whether the algorithm is the factor of the matching accuracy results. Just as the proposed method, the matching process using SIFT algorithm only extracts the features of the image without any preprocessing on the images. After we extract the feature, we continue to match two images using KNN match with the distance of 0.75, as Lowe mentioned, to find good matches of the features [25].

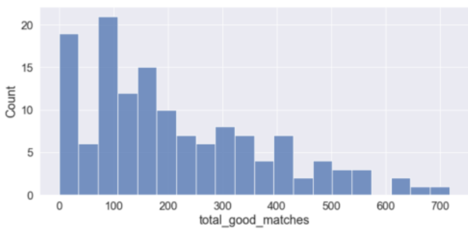

Figure 17. Distribution total matches of S1 (true positives) images using SIFT

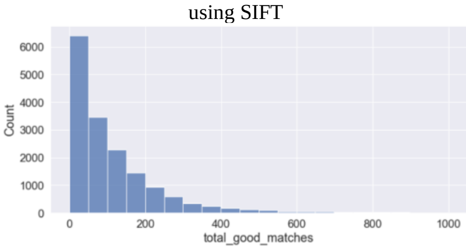

Figure 18. Distribution total matches of S2 (true negatives) images using SIFT

Figure 17 and Figure 18 are used to determine the keypoints threshold on SIFT algorithm. We are using the same method to determine the threshold and achieves the accuracy of 72.76% by thresholding the keypoints on 164 total matches. The results are better than the ORB algorithm, but we conclude that the accuracy is insufficient to match an important document identity card.

Table 4. SIFT and ORB Comparison

	Average feature extracted	Feature extraction processing time	Average matched feature	Accuracy
SIFT	**3623**	753ms	164	**72.76%**
ORB	1588	**60ms**	346	43,46%

As Table 4 shown, SIFT extracting more features but matched fewer descriptors than ORB. Extracting the image feature on SIFT does come with a processing time tradeoff. Even when SIFT is doing better than ORB, the keypoints of the images are not extracted on important areas such as name, address, card text header, face photo, or any personal information on the card image. The keypoints also matched any other parts of the images, resulting in a bias in total matches.

EKTP Important Area Feature Extracation

Feature matching on KNN-BFM will find the closest keypoint to the query image into the matched image in every area of the image. The matching result will match any feature in any image area that eventually leads to a mismatched feature or increasing the total match by matching an unnecessary area such as field name on EKTP as shown in Figure 19 and Figure 20 consecutively.

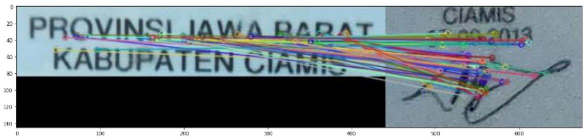

Figure 19. Example of mismatch in EKTP pair image

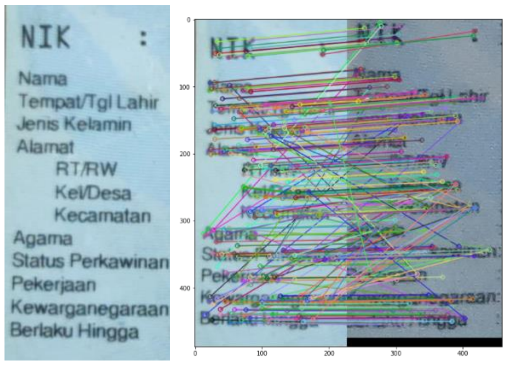

Figure 20. Field name in EKTP card image (left) and example of matching on field name on EKTP (right)

Matching on a much more important area area on EKTP will create more precise outcome in total matches. Thus, we manually segmented the 25 images dataset on the highly important area such as personal information texts (name, address, birth date, etc.) and photo faces before the matching algorithm started and continued the process the same as previous experimental results. Figure 21 shows an important segmented card area that plots the matched featured with a colored line.

Figure 21. A plot of 2 segmented photo face areas of the card image pair

As Table 5 shown, the result is improved than previous direct image matching without segmenting card important area. The result of this experiment showed that the algorithm would be much reliable if we can segment those areas into individual images, but unfortunately, in this study, such a method is not implemented to be fully automated by a machine and just segmenting the important area manually to become a proof of concept.

Not just improving the accuracy of the image matching, this will also decrease the processing time since we segment an image into a smaller chunk of images. Thus, this concept will be an interesting method to match identity documents in future studies for replacing the OCR technique in this matching case.

Table 5. High important area result image matching on 25 datasets

Field	Mean Total Matches (Threshold)	Accuracy
Citizen number	17	56.4%
Name	12	**89.6%**
Birthdate	5	54.8%
Address	12	74.5%
Face Photo	7	**86.5%**
Signature	3	43.3%

State of The Art Implementation of Verification and Matching E-KTP

To the best of our knowledge, this research is the first to implementing verification and matching tasks in this particular EKTP study case with an image pair dataset. The research result as an implementation of both verification and matching is to improve the lack of human capability on checking EKTP data conducted on previous studies [5][6] with the help of a computer (machine). A study by Kevin Akbar on how EKTP data should be collected on a repository [7] still needs a solid validation on the images uploaded to the repository conducted in our research by implementing a validation feature with verifying and matching EKTP images.

Verification task carried out with Faster R-CNN method based on the previous study by Vilás that achieved the classification of the identity document with an accuracy of 98% [8]. Even though it is not directly similar, EKTP is still an identity document, and the Faster R-CNN method consists of classification using CNN. Verification using Faster R-CNN result achieved an accuracy of 94%, 4% less than the study by Vilás. The result from Faster R-CNN is a segmented EKTP card image that will be matched on another EKTP card image. Though prior matching EKTP task can be carried out using the OCR method, based on the study by Tom Yeh, the accuracy for search and match text-based image can be improved by using the image feature itself [12]. Once again though it is not directly similar to the previous study, in our research, we tried to compare processing time and evaluation results on both OCR only and image feature only methods. The matching result shows that F-Score on OCR only, based on the study by Firhan [10], is higher than matching using image feature only.

Experimental results of this research showing many interesting points. In the verification machine learning process, the model achieves 94%, but it still detects some interesting objects in our research dataset. As Figure 10. Example of image with both false and true positive results

shown, the model results still detect rectangular shapes with text inside the object as EKTP. The model is still far from "perfect" and can be improved using a better and clean EKTP image dataset. Though we know acquiring such data is difficult, the Indonesian government can help this problem. On the matching task, using the image feature only showing its potential when the image used is segmented on the highly important area (such as name, citizen number, etc.) stated in Table 5. If this can be implemented "correctly," matching task using image feature only can be a better alternative method since processing time is improved a lot from the previous OCR only method, as shown in Table 2. And with this implementation assumption, the desired accuracy of 80% can be implemented in a real-world environment.

Both verification and matching processes show a good execution time shown in Figure 22, with an overall execution time of 2.7 seconds. Of course, the result processing time is affected by the hardware used in the research, but it shows that the method using in this research is viable to be executed in such a quick execution time.

Hopefully, the experimental result of this study can be helpful research for future Indonesian government development on EKTP business cases. The use case of this study can be a solution to Indonesia's data flow

DOI: https://doi.org/10.29207/resti.v5i4.3175

difficulty and to tried implementing a cloud-based system on confidential data such as EKTP.

```
================RESULTS================

====DETECTION USING FASTER-RCNN====
Total Cropped Image from Image 1: 2
Total Cropped Image from Image 2: 2
Execution Time: 2.5054636001586914
==================================

========MATCHING USING ORB========
--Image 1 Cropped 0 to Image 2 Cropped 0
Image 1 Total Keypoints: 2596
Image 1 Dimensions: (572, 1023, 3)
Image 2 Total Keypoints: 686
Image 2 Dimensions: (633, 1023, 3)
Total Matches: 377
Total "Good" Matches: 357
Execution Time: 0.0659644603729248 seconds
---------------------------------------------
--Image 1 Cropped 0 to Image 2 Cropped 1
Image 1 Total Keypoints: 2594
Image 1 Dimensions: (555, 992, 3)
Image 2 Total Keypoints: 945
Image 2 Dimensions: (386, 992, 3)
Total Matches: 414
Total "Good" Matches: 357
Execution Time: 0.06499791145324707 seconds
---------------------------------------------
--Image 1 Cropped 1 to Image 2 Cropped 0
Image 1 Total Keypoints: 2341
Image 1 Dimensions: (1155, 1023, 3)
Image 2 Total Keypoints: 686
Image 2 Dimensions: (633, 1023, 3)
Total Matches: 310
Total "Good" Matches: 272
Execution Time: 0.04300212860107422 seconds
---------------------------------------------
--Image 1 Cropped 1 to Image 2 Cropped 1
Image 1 Total Keypoints: 2234
Image 1 Dimensions: (1120, 992, 3)
Image 2 Total Keypoints: 945
Image 2 Dimensions: (386, 992, 3)
Total Matches: 347
Total "Good" Matches: 259
Execution Time: 0.06500029563903809 seconds
---------------------------------------------
=================================

Overall Execution Time: 2.7444283962249756 seconds
========================================
```

Figure 22. Screenshot of Combined Verification (Faster R-CNN) and Matching (ORB) Code Result.

4. Conclusion

In this study, the implementation has been carried out in both the verification and matching process. Verification process implementation was done by using the Faster R-CNN method to detect whether the EKTP was on the image dataset or not. The Faster R-CNN model achieves mAP of 94%. In the matching process, we are using the ORB algorithm to extract the image feature and KNN-BFM to find the closest keypoints on both segmented EKTP images detected by the previous verification process. The matching process using ORB only achieve an accuracy of 43.46%. Using only image features for the matching process using only image features, in this study, unfortunately, can't replace the OCR technique from the previous study. The verification detection model achieves our target of 80% mark of accuracy, but on the matching process, our study can't implement the matching process to achieve the desired target. Hence, the verification implementation succeeds in bypassing the desired target and the matching implementation; in this study, it was still far from the desired target and can't replace matching using the OCR technique.

For future studies, the matching process using the image feature shows its potential when using the correct method. Quantity and quality of the dataset, choosing other feature extraction algorithms and extracting the important area in the EKTP card image have proven to increase the accuracy for reaching the desired target. Improving the current study method for the matching process can boost the time processing from the previous OCR technique to nearly 98.6% (from 4510ms to 60ms per image) improvement.

Acknowledgment

The researchers would like to acknowledge everyone involved, including teachers, family, Lovia (https://about.lovia.life), who give us the facilities to finish this research paper, and friends for their assistance. This research would not be possible without support from them, and thanks to the Informatics, Telkom University that has given all the support to finish this research. Hopefully, this research will help other researchers and future studies. The implementation code in this paper can be accessed at the following site: https://github.com/mmhudaya-telu/ver-and-match-ektp

References

[1] S. N. Fatimah T, "Pencantuman Status Perkawinan Dalam Administrasi Perkawinan di Kantor Urusan Agama Perspektif Maqashid Syari'ah", AD, vol. 23, no. 1, pp. 79 - 92, Apr. 2020. https://doi.org/10.35719/aladalah.v23i1.28

[2] Djamhari, Eka A., et al. "Kondisi Kesejahteraan Lansia dan Perlindungan Sosial Lansia di Indonesia." Perkumpulan PRAKARSA, 2021.

[3] Putera, Roni E., and Tengku R. Valentina. "Implementasi Program KTP Elektronik (E-KTP) di Daerah Percontohan." Mimbar: Jurnal Sosial dan Pembangunan, vol. 27, no. 1, pp. 193-201. 2011 https://doi.org/10.29313/mimbar.v27i2.328

[4] Fatmasari, Fatmasari, et al. "Evaluasi Penerimaan Sistem E-ktp dengan Menggunakan Tam (Technology Acceptance Model) (Studi Kasus: Kantor Camat Ilir Timur I Palembang)." Seminar Nasional Informatika 2013, Yogyakarta, Indonesia, 2013. Universitas Pembangunan Nasional "Veteran". 2013. http://jurnal.upnyk.ac.id/index.php/semnasif/article/view/845

[5] Tawa'A, Nunung, and Andi Y. Katili. "Kemampuan Pegawai dalam Penginputan Data E-ktp di Kantor Camat Tibawa Kabupaten Gorontalo." Publik, vol. 6, no. 1, pp. 16-22. 2019 https://doi.org/10.37606/publik.v6i1.20.

[6] Iswati, Heni, and Eny Retnoningrum. "Mengukur Layanan Website E-Govqual terhadap Kepuasan Masyarakat dalam Mengakses Rekap E-KTP." Serasi, vol. 17, no. 2, pp. 101-110. 2019. https://journal.budiluhur.ac.id/index.php/serasi/article/view/949

[7] Akbar, Kevin Adhiguna, Firhan Maulana Rusli, and Hendy Irawan. "Building an ID Card Repository with Progressive Web Application to Mitigate Fraud". ArXiv. December 2020. https://arxiv.org/abs/2012.08295

[8] Vilàs Mari, P. Classification of Identity Documents Using a Deep Convolutional Neural Network. Master's Thesis, Universitat Oberta de Catalunya, Barcelona (Spain), 2018. Available online: http://hdl.handle.net/10609/73186 (accessed on December 4, 2020).

[9] Ren, S., He, K., Girshick, R.B., & Sun, J. "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal

DOI: https://doi.org/10.29207/resti.v5i4.3175

Networks". IEEE Transactions on Pattern Analysis and Machine Intelligence, 39, 1137-1149. 2015 https://doi.org/10.1109/TPAMI.2016.2577031

[10] Rusli, Firhan Maulana, Kevin Adhiguna Akbar, and Hendy Irawan. "Indonesian ID Card Extractor Using Optical Character Recognition and Natural Language Post-Processing". ArXiv, 2020. https://arxiv.org/abs/2101.05214

[11] Pratama, M. & Satyawan, Wira & Fajar, Bagus & Fikri, Rusnandi & Hamzah, Haris. "Indonesian ID Card Recognition using Convolutional Neural Networks". 2018. Proceeding of the Electrical Engineering Computer Science and Informatics. 5. https://doi.org/10.11591/eecsi.v5i5.1720

[12] Yeh, Tom & Boris Katz. "Searching documentation using text, OCR, and image". In Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval (SIGIR '09). Association for Computing Machinery, New York, NY, USA, 776–777. 2009 , pp. 83-86. 2018,https://doi.org/10.1145/1571941.1572123

[13] Karami, E., Prasad, S., & Shehata, M.. "Image Matching Using SIFT, SURF, BRIEF, and ORB: Performance Comparison for Distorted Images". ArXiv, 2017. https://arxiv.org/abs/1710.02726

[14] S. A. K. Tareen and Z. Saleem, "A comparative analysis of SIFT, SURF, KAZE, AKAZE, ORB, and BRISK," 2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), 2018, pp. 1-10 https://doi.org/10.1109/ICOMET.2018.8346440

[15] A. Jakubović and J. Velagić, "Image Feature Matching and Object Detection Using Brute-Force Matchers," 2018 International Symposium ELMAR, Z. http://doi.org/10.23919/ELMAR.2018.8534641

[16] Arlazarov, V.V., K. Bulatov, T. Chernov, and V.L. Arlazarov. "MIDV-500: a dataset for identity document analysis and recognition on mobile devices in video stream". ArXiv, 2019. https://doi.org/10.18287/2412-6179-2019-43-5-818-824

[17] Gu, Shanqing; Pednekar, Manisha; and Slater, Robert "Improve Image Classification Using Data Augmentation and Neural Networks". SMU Data Science Review: Vol. 2: No. 2, Article 1. 2019 https://scholar.smu.edu/datasciencereview/vol2/iss2/1

[18] Casado-García, Á., Domínguez, C., García-Domínguez, M. "CLoDSA: a tool for augmentation in classification, localization, detection, semantic segmentation and instance segmentation tasks". BMC Bioinformatics. 2019 https://doi.org/10.1186/s12859-019-2931-1

[19] K. Wang, B. Fang, J. Qian, S. Yang, X. Zhou, and J. Zhou. "Perspective Transformation Data Augmentation for Object Detection". IEEE Access, 2020, vol. 8, pp. 4935-4943. 2020 https://doi.org/10.1109/ACCESS.2019.2962572

[20] Srivast, Shrey & Divekar, Amit & Anilkumar, Chandu & Naik, Ishika & Kulkarni, Ved & V., Pattabiraman. "Comparative Analysis of Deep Learning Image Detection Algorithms". 2020. J Big Data 8, pp: 66. https://doi.org/10.1186/s40537-021-00434-w

[21] Wu, Yuxin, Alexander K., Francisco M., Wan-Yen L. and Ross G. "Detectron2". 2019. Facebook. Available online: https://github.com/facebookresearch/detectron2 (accessed on March 18, 2020).

[22] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," 2011 International Conference on Computer Vision, Barcelona, pp. 2564-2571. 2011, https://doi.org/10.1109/ICCV.2011.6126544

[23] Luo, Chuan & Yang, Wei & Huang, Panling & Zhou, Jun. "Overview of Image Matching Based on ORB Algorithm". 2019. Journal of Physics: Conference Series https://doi.org/10.1088/1742-6596/1237/3/032020.

[24] Henderson, Paul & Ferrari, Vittorio. "End-to-End Training of Object Class Detectors for Mean Average Precision". 2017. ACCV 2016: Computer Vision – ACCV 2016 pp 198-213. https://doi.org/10.1007/978-3-319-54193-8_13

[25] Lowe, D.G. "Distinctive Image Features from Scale-Invariant Keypoints". International Journal of Computer Vision 60, 91–110. 2004, https://doi.org/10.1023/B:VISI.0000029664.99615.94

DOI: https://doi.org/10.29207/resti.v5i4.3175