Matching Images On The Face Of A Buddha Statue Using The Scale Invariant Feature Transform (SIFT) Method

Linda Marlinda
Doctoral Program Of Computer Science
Universitas Dian Nuswantoro
Semarang, Central Java
marlinda@mhs.dinus.ac.id

Supriadi Rustad
Doctoral Program Of Computer Science
Universitas Dian Nuswantoro
Semarang, Central Java
srustad60@gmail.com

Ruri Suko Basuki
Doctoral Program Of Computer Science
Universitas Dian Nuswantoro
Semarang, Central Java
ruri@dsn.dinus.ac.id

Fikri Budiman
Doctoral Program Of Computer Science
Universitas Dian Nuswantoro
Semarang, Central Java
fikri.budiman@dsn.dinus.ac.id

Muhamad Fatchan
Informatica Engineering
Universitas Pelita Bangsa
Cikarang West Java, Indonesia
fatchan@pelitabangsa.ac.id

***Abstract*—The condition of the statue in Indonesia at this time many are missing, especially the face of the statue. 504 Buddha statues in the Borobudur temple, there are 248 statues which are now in a headless condition. The condition of the lost statue face is due to being stolen, still buried in the area around the temple, or has changed hands to another party. To find the statue's head according to the face, a matching image is needed, one of which uses the sift method. Buddha face statue matching can be done based on the characteristics of hair, eyebrow shape, lips, head shape or attributes found on the head of the statue. The image matching method can describe an image with a specific keypoint using Scale Invariant Feature Transform (SIFT), where each keypoint has a gradient orientation (GO) and gradient magnitude (GM) which are processed into features in the image registration process. The purpose of this study is to match the face images of Buddha statues with different angles. This study aims to match the face image of the original statue with the image of the statue chosen from different angles. The results of this study are that the same image with different angles obtains 90% accuracy.**

Keywords—face detection, face matching, SIFT, Statue Buddha

I. INTRODUCTION

In the Borobudur temple in Indonesia, there are several Buddha statue, with the condition of the face of the statue being damaged and a normal statue [1][2][3]. This buddha statue has the characteristics of high relief art, naturalist style, depictions of fabric folds (drapes), the sitting attitude of figures, and the presence of headdress ribbons. On the face of the Buddha statues, there are characteristics including lip patterns, at the tip of the head there is curly hair of the Buddha and always clockwise and bun (ushnisa), forehead on the face there is a small protrusion called (urna), on the neck, there are three that symbolize patience. The long ears show maha. The tip of the nose and eyes half-closed shows concentration[4][5].

Buddha statue found in temples in Indonesia must be preserved and labeled or informed so that future generations can get to know and learn history. To overcome the above problems, it is proposed to detect image statues and matching using the Sift method so that the statues data can be simulated and digitally[[6][7].

The scale-invariant feature transformation method (SIFT) in matching image statues has matching values that can be influenced by parameters including, angle of view, scale changes, size, rotation and light intensity in the image[8][9] so that the effectiveness value can be optimal if it is directly. Proportional to the image recognition according to the number pixel. Before going to the matching process, the image of the statue must be detected by looking for features that distinguish it from other statues. In the detection process will fail if the statue image changes in size, lighting, rotation and shooting angle[9][6][10].

II. RELATED WOKS

A. Statue buddha in Indonesia

At present, the condition of the statue in the Borobudur temple in Indonesia has a headless condition and there are some missing ones. indeed lost since before being restored, lost due to being stolen, or taken by another party, there are some abnormal or damaged statuary conditions, without a head or missing some other body parts totaling 248 [3][5][2][6].

From these problems, a matching statue is needed to find the missing statue head. Each statues head has different characteristics that can be seen from the hair, eyebrow shape, lips, and different head shapes. the facial expressions of the statue of the Buddha's head are on average the same, only the hand positions are different, there are the positions of abaya mudra, mudra, and there are turning chakra[11][4][3].

B. Maintaining Image Maching

In image matching, measuring the similarity between two images is very important. Each image is made into quantitative characteristics that can be measured to be smaller so that it is called image features, and the process of comparing image features with other images is called image matching. Feature-based techniques are used in image matching. This technique can clean or find image features such as corners, edges, and blobs. The formation process determines the number of points in the image of a statue that has a complicated background condition.

1. Feature Detection, the process of locating positions in Buddhist image images, visual features described in the image region that contain structural information, such as

patterns, edges, and angles. This detection can search across all pixels of the image and traverse various sizes of the image area. Scale-invariant detection is a picture of a different statue that has a different scale and can detect the same region. The choice of pixel size makes the detection process in features for two different images and have the same background. Many features can be extracted from digital images including the edge is the boundary to mark between different fields in the picture, for example, different brightness levels, or different textures. Angles are found at the apex in the correlation function or the point where the edges intersect[9][12][13].

2. Feature Description, Describe locations in the form of histogram measurements of Buddha statue images. In describing the image, the level of noise and changes in the angle of view and imaging of the metric conditions must be considered in a piece of concise information about the color, texture, and shape descriptors[13][14][15][16].
3. Feature matching, in the process of detection and description of image features, is represented as histograms, each image must be abstracted as a set of local features. matching two images, we need a search technique that compares each pair of features from each image based on a measure of similarity (Euclidian, or Mahalanobis). So the feature matching procedure consists of three parts: similarity measures, matching strategies, and search techniques[17][18][19].

C. *SIFT Algorithm*

Using the Scale Invariant Feature Transform (SIFT) method, which is proposed to overcome the above problem, this algorithm can detect similar feature points in each available image and then illustrate these points with feature vectors that do not change for scale and rotation, and some do not change in illumination and change in perspective. The use of SIFT algorithms with different points of view must have similar feature vectors [20][19]. This method has two process parts, namely the extraction matching feature SIFT and the matching feature SIFT. Extraction involves finding and describing the interests of a region or point, whereas matching means finding correspondence between features in different images. SIFT method is a matching method for images that have similarities with other similar images [20][21]

III. Research Methodology

In this section, we present the stages of a research methodology for detecting and matching statues, in Figure 1 as follows:

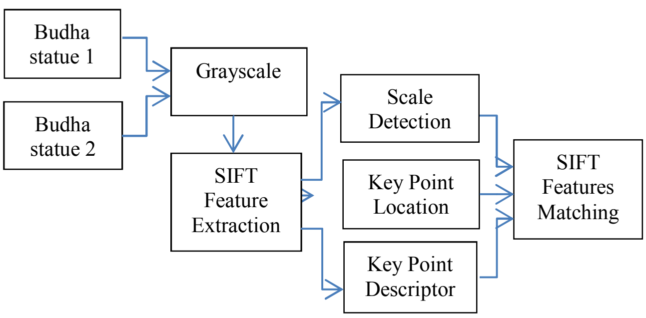

Fig. 1. Extraction and matching of SIFT buddha statues

The SIFT algorithm extracts key points that do not change for scale and rotation using difference images that have different scales. Rotation of each invariant is achieved for each keypoint location based on the gradient direction of the statue image including:[22][23][24].

The Gaussian weighting function with s equal to one half the width of the descriptor window is used to assign weights to the magnitude of each sample point. SIFT descriptors that are invariant to scaling, rotation, and transformation will be calculated using the following four main steps:[14][25][18]

1. Scale Space Extrema Detection[15][26][27][28]

The function, L (x, y, s) is defined as the scale of the image generated by the convolution of the Gaussian function, G (x, y, s) and the input image I (x, y):

$$L(x,y,\sigma) = G(x,y,\sigma) * I(x,y) \quad (1)$$

$$G(x,y,\sigma) = \frac{1}{2\pi\sigma^2} e - (x^2 + y^2)/2\sigma^2 \quad (2)$$

Where convolution, (x, y) is the pixel coordinate and s is the scale space factor or variant of the normal Gaussian distribution. For efficient detection of a stable and reliable keypoint, the DOG (Difference Of Gaussian) function, distances are calculated by combining the differences of two nearby scales separated by a constant scaling factor 'k' with the input image

$$D(x,y,\sigma) = (G(x,y,k\sigma) - G(x,y,\sigma)) * I(x,y)$$
$$= L(x,y,\sigma) - L(x,y,\sigma) \quad (3)$$

2. Keypoint Localization

Taylor expansion of scale-space function so the sample points are:

$$D(x) = D + \frac{\partial D^T}{\partial x} x + \frac{1}{2} x^T \frac{\partial^2 D}{\partial x^2} x \quad (4)$$

To determine the location of the extreme, the derivative of w.r.t. taken and set to zero.

$$\text{ɤ} = -\frac{\partial^2 D^{-1}}{\partial x^2} \frac{\partial D}{\partial x} \quad (5)$$

3. Orientation Assignment

To achieve rotational invariance, each keypoint is given an orientation. For each sample of the Gaussian smoothed image, $L\ (x,\ y)$, the magnitude of the gradient, $m\ (x,\ y)$, and the correlation, 0 $(x,\ y)$ is calculated by pixel difference[27]:

$$m(x,y) = \sqrt{(L(x+1,y) - L(x-1y))^2 + (L(x,y+1) - L(x,y-1))^2} \quad (6)$$

$$\text{Ø}(x,y) = tan^{-1} ((L(x,y+1) - L(x,y-1)) / (L(x+1,y) - (x-1,y))) \quad (7)$$

The direction of the feature's gradient is calculated using a gradient oriented histogram. The peak histogram orientation is the dominant direction of the local gradient.

4. Keypoint Descriptor Generation

The orientation histogram value, with an array of $4x4$ from the histogram and 8 orientation produces $4x4x8$ = 128. The feature vector is modified to reduce the effects of illumination change. The vector is normalized to unit length. Changes in image contrast where each pixel value multiplied by a constant will increase the gradient with the same constant, so this contrast change will be canceled by vector normalization. the descriptors are invariant for affine changes in illumination. However, changes in non-linear illumination can also occur due to camera saturation or due to changes in illumination that affect 3D surfaces with different orientations with different amounts. This effect can cause large changes in relative magnitudes for some gradients but tends to affect the gradient orientation

IV. Result And Discussion

In this research the implementation of input data will be processed in the form of two image sample files in JPG / JPEG format, the source of the image is obtained from the data set used by alicevision.github.io. in this case there are two class labels, namely the face of the original statue and the face of the dataset statue.

Fig. 2. Buddha statue image 1 and statue image 2

Explain about the histogram of image 1 which has a pixel value and an average value per RGB color, Buddha statue image 1 has a total of pixels on the histogram 4,213,440 with gray 121.66, red 124.04, green 121.23 and blue 117.95. Buddha statue image 2 has a total pixel total pixels on the 4,213,440 histogram with gray 124.19, red 128.64, green 123.33, and blue 116.95.

Keypoint search uses Gaussian filter calculation by multiplying the image value and producing a convolutional value. After that, the Difference of Gaussian calculation is done. Taking the maximum or minimum value of the Difference of Gaussian from the Keypoint, if there are many key points, then low contrast must be eliminated. The process of removing a key point with low quality can be seen in the maximum area and the minimum area, if the value is less than the threshold then the point does not become a key point. the result of maching point with sift feature extraction on two images with the same engel and different background.

Fig. 3. Features extraction of buddha image

The result of feature extraction of Buddhist images given a threshold value in the database in Figure 3 has 364 x 128 image features with a scale of 909 x 1, while a Buddha statue image 2 has 311 x 128 image features with a scale of 848 x 1. In Figure 3 you can see features other than imagery extracted buddha, feature values are not counted in the image database. Look for a statue image match, after having the feature and descriptor information, the system can make a matching feature accordingly. Descriptors in each image database are compared with descriptors in background images on the table, which are defined as keypoints with a minimum value of Euclidean distances in the descriptor space. After the matching stage, it is further filtered to remove features in the image database that are matched with one feature in the background image on the table which is the result of the feature matching stage, and still looks outlier or false match features an inlier or feature matches correct (true match).

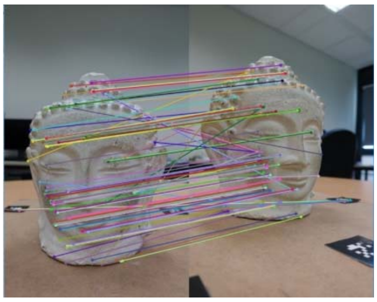

Fig. 4. Matching point sift method for buddha statue

A. Test Threshold Value

The level of accuracy is defined when all images are correctly detected, regardless of the selected threshold value. For example, if the image of a background table, the results will be considered accurate when the top 2 database images match with Buddha in the background table. It is said to fail when the image database is not detected correctly in the background table image but is in the top two image databases. Next, the threshold value used to determine the average threshold value corresponds to the lowest number of matches of the successfully detected feature. For example, the background table image with image Buddha 2 with the average value of the feature matching threshold is 80.28. Table 1 shows the accuracy of the algorithm as a function to determine figure 1 with figure 2.

TABLE I. THE ACCURACY OF THE ALGORITHM

Number of Images	Total Pixels	image features	Level of accuracy
Image Buddha statue 1	4213440	364 x 128	90
Image Buddha statue 2	4213440	311 x 128	90

Test results on two Buddhist images show that Buddha images can be recognized 90%. With different angles and different backgrounds, the level of background imagery has decreased inaccuracy at medium and long distances. the reason for the decline is the increase in the image of Buddha in the background image. With the image taken is taken at a closer distance. Therefore, image background which has a little image has a large sack on the foreground This results in a far more accurate result because the outlier pairs of the image database that are not in the possibility of filtering are higher because of the large boundary pixel area. Conversely, the image of the background table with more images taken at a greater distance to capture the image of the whole culture, and will happen. The possibility of filtering the wrong features will decrease because there are more background pixels in the image.

V. CONCLUSION AND FUTURE WORKS

In the comparison of the two images compared to produce a histogram value with several different pixels, the histogram value is used as a reference as an additional parameter to determine the similarity of the location of the keypoint and the number of pixels in the Buddha statue image. Based on the test results obtained accuracy in identifying images in digital images using SIFT (Scale Invariant Feature Transform) Algorithm, with a level of accuracy of 90% for normal and non-deformed Buddha statue faces.

The object used in the next research is the damaged Buddha statue face object which will be identified and matched with a normal Buddha statue face image.

REFERENCES

[1] R. Agung Sugihartono, D. Dharsono, G. Guntur, and M. Rusnoto Susanto, "Therianthropic character in Garuda statue and relief," *Int. J. Recent Technol. Eng.*, vol. 8, no. 1C2, pp. 679–683, 2019.

[2] B. Renoust *et al.*, "Buda.art: A multimodal content-based analysis and retrieval system for Buddha statues," *MM 2019 - Proc. 27th ACM Int. Conf. Multimed.*, pp. 1062–1064, 2019.

[3] H. Setyawan *et al.*, "Matching Study Of The Head Of Arca Buddha Borobudur Temple Stage II Hari Setyawan, Agus Hendratno, Marsis Sutopo, Jati Kurniawan, Puji Santosa, Irawan Setiyawan," vol. XI, pp. 4–24, 2017.

[4] A. Gruen, F. Remondino, and L. Zhang, "Automated modeling of the great Buddha Statue in Bamiyan, Afghanistan," *Int. Arch.*, 2003.

[5] A. Y. Rahman and S. Sumpeno, "Segmentation of Arca at Mpu Tantular Sidoarjo Museum Using Learning Vector Quantization," no. July 2017, pp. 16–21, 2016.

[6] B. Renoust *et al.*, "Historical and Modern Features for Buddha Statue Classification," pp. 23–30, 2019.

[7] S. Temple, *Relief and Statue*. .

[8] O. Recognition, "Image Matching."

[9] X. Sun, P. Wu, and S. C. H. Hoi, "Face detection using deep learning: An improved faster RCNN approach," *Neurocomputing*, vol. 299, pp. 42–50, 2018.

[10] M. I. Ouloul, Z. Moutakki, K. Afdel, and A. Amghar, "An efficient face recognition using SIFT descriptor in RGB-D images," *Int. J. Electr. Comput. Eng.*, vol. 5, no. 6, pp. 1227–1233, 2015.

[11] B. Renoust *et al.*, "Historical and modern features for Buddha statue classification," *SUMAC 2019 - Proc. 1st Work. Struct. Underst. Multimed. Herit. Contents, co-located with MM 2019*, pp. 23–30, 2019.

[12] H. Wang, "SPCDet : Enhancing Object Detection with Combined Feature Fusing," no. 2017, pp. 236–251, 2019.

[13] Y. Li, S. Wang, Q. Tian, and X. Ding, "A survey of recent advances in visual feature detection," *Neurocomputing*, vol. 149, no. PB, pp. 736–751, 2015.

[14] G. Minster, G. Moreau, and H. Saito, "Geolocation for printed maps using line segment-based SIFT-like feature matching," *Proc. 2015 IEEE Int. Symp. Mix. Augment. Real. Work. ISMARW 2015*, pp. 88–93, 2015.

[15] X. Wang *et al.*, "Feature Fusion of HOG and WLD for Facial Expression Recognition," pp. 227–232, 2013.

[16] D. Metusala, Fauziah, D. A. Lestari, J. Damaiyani, S. Mas'udah, and H. Setyawan, "The identification of plant reliefs in the lalitavistara story of Borobudur temple, central Java, Indonesia," *Biodiversitas*, vol. 21, no. 5, pp. 2206–2215, 2020.

[17] P. Mandle, "A Survey : Image Matching," vol. 4, no. 1, pp. 4–7, 2015.

[18] A. D. Narhare and G. V. Molke, "Trademark detection using SIFT features matching," *Proc. - 1st Int. Conf. Comput. Commun. Control Autom. ICCUBEA 2015*, pp. 684–688, 2015.

[19] C. K. Toth, H. Ju, and D. A. Grejner-Brzezinska, "Experiences with using SIFT for multiple image domain matching," *Am. Soc. Photogramm. Remote Sens. Annu. Conf. 2010 Oppor. Emerg. Geospatial Technol.*, vol. 2, pp. 938–945, 2010.

[20] D. Papp, F. Mogyorósi, and G. Szucs, "Image matching for individual recognition with SIFT, RANSAC and MCL," *CEUR Workshop Proc.*, vol. 1866, no. 1, 2017.

[21] S. Ansari, "A Review on SIFT and SURF for Underwater Image Feature Detection and Matching," *Proc. 2019 3rd IEEE Int. Conf. Electr. Comput. Commun. Technol. ICECCT 2019*, pp. 1–4, 2019.

[22] I. Ilievski, T. Akhtar, J. Feng, and C. A. Shoemaker, "Efficient hyperparameter optimization of deep learning algorithms using deterministic RBF surrogates," *31st AAAI Conf. Artif. Intell. AAAI 2017*, pp. 822–829, 2017.

[23] R. Tsuzuki, H. Zen, K. Tokuda, T. Kitamura, M. Bulut, and S. S. Narayanan, "Constructing emotional speech synthesizers with limited speech database," *8th Int. Conf. Spok. Lang. Process. ICSLP 2004*, pp. 1185–1188, 2004.

[24] J. Ma, X. Jiang, J. Jiang, and Y. Gao, "Feature-guided Gaussian mixture model for image matching," *Pattern Recognit.*, vol. 92, pp. 231–245, 2019.

[25] G. Guo and A. Lai, "A survey on still image based human action recognition," *Pattern Recognit.*, vol. 47, no. 10, pp. 3343–3361, 2014.

[26] J. Sarazyn, "(12) Patent Application Publication (10) Pub . No .: US 2014 / 0332121 A1 Eiongation (%) Patent Application Publication," vol. 1, no. 19, p. 33, 2014.

[27] A. Y. Rahman, S. Sumpeno, and M. H. Purnomo, "Arca detection and matching using Scale Invariant Feature Transform (SIFT) method of stereo camera," *Proc. - 2017 Int. Conf. Soft Comput. Intell. Syst. Inf. Technol. Build. Intell. Through IOT Big Data, ICSIIT 2017*, vol. 2018-Janua, pp. 66–71, 2017.

[28] A. Setiyawan and S. Basuki, "Image Matching based on Scale Invariant Feature Transform (SIFT) using Arc Cosinus," *J. Tek. Inform. UDINUS*, 1999.