

# Superresolution Reconstruction in Automatic Thai Sign Language Feature Extraction Using Adaptive Triangulation Interpolation<sup>1</sup>

<https://doi.org/10.3991/ijoe.v18i02.28147>

Eakbodin Gedkhaw, Mahasak Ketcham<sup>(✉)</sup>  
King Mongkut's University of Technology North Bangkok, Bangkok, Thailand  
mahasak.k@itd.kmutnb.ac.th

**Abstract**—Superresolution is an image processing technique for improving image quality and enhancing low-resolution images. This paper presents a novel interpolation method for increasing superresolution reconstruction effectiveness using a triangulation interpolation algorithm for automatic Thai sign language feature extraction. This approach uses three neighboring pixels for estimation. The experiment compared the superresolution reconstruction performance using the triangulation interpolation algorithm and the nearest-neighbor, SRCNN, bilinear, bicubic, GPR, and NEDI methods. The superresolution reconstruction using the triangulation-improved interpolation technique provided the best PSNR measurements of image quality between the original and superresolution-reconstructed images. The PSNR value of the sign language image was 40.608, improving performance by 13.15%. The Thai sign language gesture recognition using 2D convolutional neural networks showed that the designed model increased the gesture recognition effectiveness with an accuracy of 0.95 and loss of 0.14. Thus, this study provides state-of-the-art superresolution reconstruction for automatic Thai sign language gesture recognition.

**Keywords**—superresolution, triangulation interpolation, Thai sign language

## 1 Introduction

Sign language is the communication way among people who are hearing impaired. On the other hand, most normal people cannot understand sign language and it's a major problem to communicate between deaf and normal. It's challenging to discover how to communicate between different people in easy, researcher use superresolution technique in this paper. Superresolution (SR) is an image processing technique for improving image quality and enhancing low-resolution (LR) images into high-resolution (HR) images [1]. SR was first studied by Tsai et al. in 1980. In the last few years, this technique has been studied for various purposes, including remote detection, object detection, and recognition [2, 3]. This technique has also been applied to

---

<sup>1</sup> The authors hereby confirm that they have obtained the consent of the persons depicted in the photographs for publication.

aerial photos, spatial images, satellite images [5], and medical applications such as X-ray imaging. Recently, SR has been classified as a reconstruction-based method, an interpolation-based method, and an example learning-based method [1]. The SR reconstruction process can be classified as multi-frame and single-frame [2]. Both reconstruction techniques improve and enhance original single images and scale and rotate to produce higher-resolution images. This study focused on a novel method of scaling up and zooming using triangulation interpolation to produce SR images.

SR has used interpolation techniques such as nearest-neighbor, bilinear, and bicubic, which are not complicated. However, these techniques produce jagged edges and blurry effects because the high-frequency image components are lost [1]. Li and Orchard [6] suggested edge-directed interpolation that improved SR by using local variance and improved variance estimation of a HR image from a LR image based on geometric duality. Jing and Wu [7] suggested fast interpolation using the weighted inverse distance. Yang et al. [8] proposed a sparse representation model (SRM) that reconstructed HR images by applying the nonlocal autoregressive model to scale down the LR image. Dong et al. [9] presented a random forest and convolutional neural network (CNN), which introduced a novel SR reconstruction method. Later, Timofte et al. [10] presented a superresolution convolutional neural network (SRCNN), which used a three-step CNN to reconstruct a new LR image but with higher resolution. However, the process was complicated and more time-consuming than the linear conventional method.

Sign language uses hand signals or other body parts to communicate [11, 12] among the hearing impaired. Sign language recognition studies have become popular among computer vision and cognitive science researchers. Computer training has been challenging when attempting to enable computers to understand sign language gestures. Improvements in video trimming, sign extraction, sign video background modeling, sign feature representation, and sign classification [13] are required to improve sign language interpretation for general public use. Human language interpretation for computer understanding remains limited; hand tracking, finger shape modeling, torso tracking, head movements, and pattern recognition [14] are complicated and difficult for real-time sign language recognition [15].

Two methods that are frequently used for sign language recognition are image-based and sensor-based methods [16]. While the sensor-based method requires wearable electronic equipment such as a data glove [17] and EMG, the image-based method has three steps. In the first step, cameras or sensor devices receive images. Digital camera images require preprocessing before transfer to the second step. Data from sensor devices can be transferred directly to the second step without preprocessing. In the next step, features from received data or images are identified and extracted [18]. Recognition is performed in this step, which distinguishes gestures of the hands and other body parts. The computer perceives the meaning of these gestures or human language. In hand gesture recognition studies from digital camera images, such as Kinect [19] or leap motion [20], the first image capture step is crucial. Images must be clear (crisp), clean, and sharp for successful feature extraction. LR and small images limit the feature extraction process. Therefore, LR images are scaled up and reconstructed to HR images for feature extraction.

This study offers a novel interpolation using the triangulation editing technique. An unknown HR image pixel was interpolated within the interval calculated from the nearest three triangular points. The interpolation weight calculation using triangulation interpolation minimized the expected error variance and extracted features for automatic sign language recognition. We discuss and compare the abilities of theoretical methods and algorithms to reduce blur and blocking and enhance resolution. These discussions establish a basis for future SR algorithm improvements for feature extraction in automatic sign language gesture recognition. In this paper, we present a SR algorithm analysis and evaluate the process results.

## 2 Background and notation

### 2.1 Single-image superresolution

SR transforms multi-frame and single-frame LR images into HR images. This technique resolves many issues related to remote sensing, robot vision, video enhancement, and medical imaging [21]. These processes enlarge the input to enhance the image resolution, increase the density, and reduce noise. The reconstructed image results in higher precision, as shown in Figure 1. The SR image has low noise and reduced blur because the spatial frequency is maximized. These properties are essential for image fusion [22].

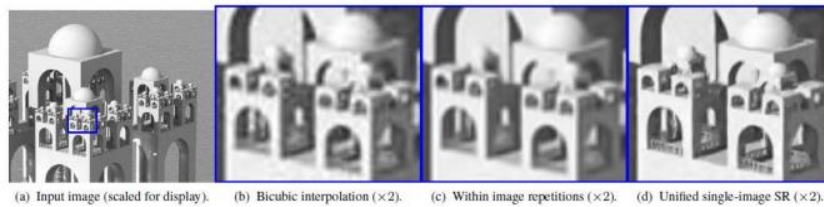


Fig. 1. SR technique [23]

There are three common SR-type methods: interpolation-based methods, reconstruction-based methods, and learning-based methods [5, 26, 27]. In general, SR assesses unknown HR pixels using neighborhood pixels with reconstruction properties. The interpolation-based method is the least complicated for enhancing image spatial resolution. We can assume that this method creates a continuous function from the pixel in the LR image, after which this function is resampled and reconstructed using techniques such as linear, bicubic, B-spline kernels, Lanczos unsampling, and new edge-directed interpolation (NEDI) [24]. Although these methods can effectively reconstruct the HR image, the jagged edge still presents a blurry effect. The reconstruction-based method uses signal processing technology to prevent the image reversing system and restore the high-frequency data loss that occurred during the shooting. The HR image can be restored with sharp edges. However, high-frequency textured area details are commonly lost. This method uses techniques such as maxi-

mum a posteriori (MAP), projection onto convex sets (POCS), and regularization [24]. This last method, which is learning based, is commonly used for image SR. The mapping relationship between LR and HR images is learned from the double patched LR/HR sample dataset to predict the missing LR image details. Example-based and dictionary-based techniques are classified under this method.

Researchers have offered solutions to image SR from a single-framed LR image by learning from the enhancement process. In single-framed image SR, purpose HR image reconstruction results are estimated from LR images, typically considered to be the inaccessible reduced version, and reconstruct HR images related to LR images, as described in the following equation [25].

$$Y = SHX + n \quad (1)$$

Where Y is the HR image from the LR image, S is the sampling degradation matrix, H is the blurring synthesis operator, X is the corresponding HR image, and n is additive white Gaussian noise.

## 2.2 Interpolation method

For the single-frame image SR using the interpolation-based method, the nearest pixel is used to interpolate the editing pixel. This method is straightforward without complicated calculations. The interpolation-based method reconstructs HR images using the edge radius and robustness. This method estimates or interpolates the data of an empty pixel in a raster from the neighboring pixel to resolve problems when the data are insufficient and predicts an unknown value from any pixel in the image. A raster or bitmap image is an image composed of small colored dots known as pixels arranged together in a grid format. Red, green, and blue (RGB) is the most common color dot standard with a fixed number.

The raster image is based on each pixel's color reading. Each pixel has a specific color value at each position and is suitable for images requiring coloring since color generation and resolution determination are quite easy. However, to scale up the image, the resolution decreases since the number of pixels is limited, and resolution enhancement increases the file size.

According to Figure 2, the input is a pixel placed at the center of the cell that might be different from the neighboring points. When creating the raster with the interpolation method, the result can be reduced to the created area boundary size. When the data point is placed inside the cell, it cannot be guaranteed that the cells correctly match the original values. The interpolation method approximates the missing value based on the neighboring object distribution. We assume that the results are similar.

The interpolation method offers several SR image processes, all of which use point data as an input. It can be used in inverse distance weighting [26], natural neighbors [27] or kriging [1] to generate images via the 3D analyst user interface. The reconstruction process, Trend, requires programming or instruction to interpolate the value. Each method assumes the estimated value and assigns it to a particular cell based on the data simulation and data distribution. The difference method may yield different results.

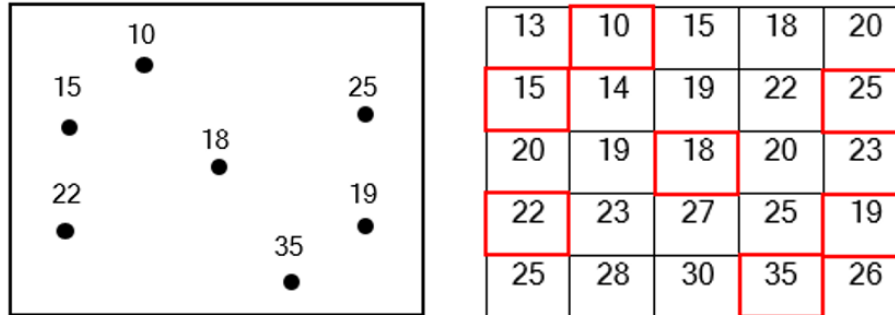


Fig. 2. Pixel interpolation

The inverse distance weighting [26] interpolates values by randomly selecting sampling points that affect the target cell, and the effect gradually decreases according to the distance between the target cell and the sample. Thus, the point located nearest to the target cell is calculated with a weight greater than the farther point. The number of cells or points within the radius can be specified for the calculation. This method is suitable if the variables used to create the map are adjusted according to the distance from the sample point.

The natural neighbor [27] interpolation method relies on the average weight of the obtained data. However, other than interpolating values using all neighboring points and weighing gradually, the natural neighbor method creates a Delauney triangle for the sample point and selects a point near the interpolated area vertex. The value weight is proportional to that area. This method is suitable for reconstructing images from a sample point with an uncertain density distribution. This interpolation method can be adapted to images that do not require radially adjusted variables, numbers, or weights.

Kriging [1] interpolation assumes a value from the distance or direction between each sample point with a spatial relationship that describes image changes. The kriging method selects a mathematical equation suitable for the selected sample point and all sample points within a given radius to calculate the result. This method combines statistical analysis, variogram modeling, and image reconstruction with image variance monitoring. This method is often used when distance or direction affects the change in information and is often used in agronomy and geology.

### 2.3 Triangulation interpolation

The triangulation interpolation method studies the relationship between the function and data by using two types of data—namely, the actual data and error terms—to implement a function. The actual data result in a function data point directly related to the data and are easy to implement. The error term is converted to a function corresponding to the existing data. It can be used to represent a relationship or forecast using certain rules, it can use that information to derive potential data trends, or it can implement a function from experimental data.

This interpolation method, which is a technique for finding a range of real data points, writes a function that passes every data point to estimate or forecast various values between points in the desired range. This technique is suitable for data points that are known or highly accurate and obtained through complex calculations or highly reliable experimental data.

We present a calculation method using triangulation to accurately estimate or forecast values in SR imagery. Triangular interpolation enlarges an image by magnifying the space between the colored dots. When enlarging an image even more, the empty pixels are even larger, as shown in Figure, 3, resulting in a lack of detail and image clarity and blurring the image. LR images can be converted to a Delaunay triangle, and the values of those empty pixels are calculated by finding the correlation of nearby pixel points using three pixel dots, as shown in Figure 3.

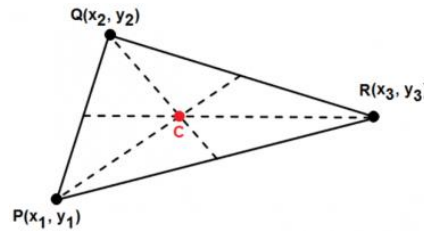


Fig. 3. Relationship among neighborhood pixels

In Figure 3, point C can be calculated from the three pixels using equation 2.

Given the coordinates of the three vertices of a triangle  $P(x_1, y_1)$ ,  $Q(x_2, y_2)$  and  $R(x_3, y_3)$ ,

$$C = \left( \frac{x_1+x_2+x_3}{3}, \frac{y_1+y_2+y_3}{3} \right) \quad (2)$$

The area of the triangle (T) is given by equation 3:

$$T = \frac{1}{2} |(x_1y_2 + x_2y_3 + x_3y_1) - (x_2y_1 + x_3y_2 + x_1y_3)| \quad (3)$$

To understand the information distribution from the finite number of P, R, and Q points in Figure 3, the data at point C are evaluated or interpolated from the data at different points nearby. For example, temperature and air pressure distributions are observed in this way. In repeated observations, values at point C are evaluated in the P, R, and Q data models, which contain observational data at random positions. Data at the three observations nearest to C provide an appropriate and effective data value near the linearly interpolated observation point.

In determining the three nearest points for C, all points in that field area are divided into triangles. In other words, data fields are partitioned into a series of triangles. A triangle numerically estimated in the elongated Delaunay net shape is known as the optimal triangular partition or Delaunay triangulation. This technique can be applied to any triangle. There are many algorithms for generating Delaunay triangulation, which are classified into two types. A static algorithm implicitly assumes that all data

points are known a priori. Dynamic algorithms allow effective triangles to be updated by inserting new points.

## 2.4 Thai sign language

Thai sign language (TSL) uses hand gestures, hand positions, and facial expressions to communicate with hearing-impaired people [6]. Therefore, TSL is important for people with impaired hearing. There are five important sign language elements.

1. Hand shape: Various gestures such as spreading hands and fists may be similar, but they have different meanings in different positions.
2. Location: The hand position should be at an easily visible and clear level, such as at the head, near the face, and not below the waist. The same hand posture in different positions conveys different meanings.
3. Movement: Body, face, and hand movement imparts different meanings.
4. Palm Orientation: Hand and palm movement convey meaning depending on direction. The hand shape may be the same, but different palm orientations give different meanings.
5. Facial Expression: Expressions, such as raising eyebrows, inflating the cheeks, and expressions of surprise are important gestures. Figure 4 shows examples.



Fig. 4. Sign language expressions [28]

## 2.5 Sign language recognition

Sign language recognition enables computers to recognize gesture meaning. Researchers have studied various approaches to find methods or processes for enabling computers to interact with humans. Examples include using the support vector machine to recognize the fixed spelling of one finger in American Sign Language, dynamic hand gesture recognition from spiral hand gestures [29], and American character recognition using Microsoft Kinect [30]. In general, the first stage of gesture recognition is primary data processing, or preprocessing. Preprocessing imports gesture data into the system. Then, the information goes through the gesture feature extraction process to select the unique and special features to discern gesture meaning [31]. Next, the information undergoes identification, or classification, to recognize hand gestures. Figure 5 outlines the general process.



**Fig. 5.** Gesture feature recognition process

## 2.6 Related research

Feng et al. [32] proposed a method for reconstructing HR images through learning. By sending LR images to the network and learning the nonlinear mapping relationship between LR and HR images over the network, they found that several HR levels combined image properties with different scales. Rough nonlinear mapping between LR and HR images is followed by detailed mapping. Unified HR image reconstruction loses some LR image properties and adversely affects the quality of the newly created image. Multilevel HR image reconstruction can resolve LR image imperfections. HR images reconstructed by multilevel HR networks and true HR images are combined by image guide filters to create new images. The newly created images are used for training in the second stage of the reconstruction network, creating multilevel ultrahigh-resolution images. The newly created image effectively compensates for detail and texture information lost in LR images and improves HR image reproduction accuracy and speed.

Huang et al. [2] introduced a method for reconstructing a single SR image through sparse representation, which has become widespread. By utilizing robust edge-preserving smoothing superresolution (REPS-SR) in the sparse region, EPS standardization uses domain gradient filtering to preserve edges and reduce spots. In addition, smoothness factors determined by estimating the level of LR image noise without manual interference optimize balance between the precision ranges. The proposed EPS standardization offers a recoil algorithm for SR images. For LR images, the proposed adaptive smoothing makes our method effective at different noise levels. The results indicated that the proposed method was able to preserve image edges and reduce noise and performed better than current modern methods for noise-sensitive images.

Zhang and Wu [1] presented image SR using windowed ordinary kriging interpolation, which is widely used in geoscience. The proposed method accounts for the pixel data intensity and geometry distance. A new distance-intensity definition and a window in the area surrounding each unknown HR pixel were used to implement the algorithm. This approach improved weight and edge maintenance. The results demonstrated effective peak signal-to-noise ratio (PSNR) and visual perception in the proposed method.

Chen et al. [4] used sequential gradient constrained regression for single-image superresolution (SGCRSR) to present single-image superresolution (SISR), which produced HR images and improved LR single image quality. In particular, the regression-based SISR approach learns the mapping between the LR and HR double patches. However, the advanced edited images created by this method are often affected by visual artifacts, as no additional restrictions are enforced. To mitigate these shortcom-



ings, the researcher proposed that SGCRSR use continuous sequential regression, which provides an efficient way to integrate basic learning methods and reconstruction. Flaws in the training and testing procedure were identified to offer improved A + (EA +). Then, the EA + model was applied in the dual intensity-gradient domain to create the superresolution gradient constrained regression (GCR). Finally, SGCRSR was created to gradually improve the edited image quality. In the experiments, SGCRSR performed better than the SR algorithms.

Zhang et al. [21] used image SR via a novel cascaded CNN framework. Their method achieved excellent self-deep learning capabilities in three steps: feature extraction, detail forecasting, and reconstruction. The feature extraction process offers feature mapping to restore LR image features. In the detail forecasting process, the missing original LR image details are predicted across the entire network. Finally, learning produces SR images by overlaying the missing details and the original LR image to avoid gradient explosions. Gradients are used to train the proposed CNN framework.

Changa et al. [25] used SISR with collaborative representation and nonlocal self-similarity. A SISR algorithm was proposed using synergistic representation and the similarity of local nature images. The representative collaboration method creates a manual projection matrix from a training set of HR images, and then the trained self-projection matrix creates a clustering and collaborative representation (CRR) model to introduce external human resources. In addition, human resource images are internally exploited before being considered to guarantee reliable estimation in a nonlocal low-rank regularization (NLR) format. As the CRR term and the NLR term are complementary, they are merged to eventually create a new framework for SR recovery that implements an iterative algorithm to gradually improve quality. Comprehensive experimental results showed that the proposed method was capable of delivering higher-quality SR results than many modern SR methods.

Enríquez-Cervantes and Rodríguez-Dagnino [27] presented SR visualization algorithms. They performed SR image reconstruction using natural neighbor interpolation. The algorithm is divided into two steps: image registration and SR color image reproduction. In the first step, when image shifting is unknown, the algorithm calculates these estimations by solving systems of linear equations proposed by Brada et al. After that, the LR image pixels are mapped to the HR table by recalculating the coordinates using a moving vector. The corresponding pixel values with the HR grids are interpolated using natural interpolation with neighborhood interpolation, which is a weighted average interpolation method for data scattered by neighboring pixel Voronoi polygon areas.

### **3 Proposed methodology**

In this study, we present SR reconstruction using triangulation interpolation algorithms for feature extraction in automatic sign language gesture recognition. Experiments evaluated the performance of the method by measuring errors and comparing the SR image with the original image using triangulation interpolation. The images

used in both tests were based on the same set of tests. The objective of this experiment was to evaluate the effectiveness of the triangulation interpolation technique algorithm, which focused on efficiency in reducing image blurring and edge blocking. It also evaluated whether it could create a corner in the form of a square edge and a more detailed image by comparing SR image reconstruction errors with the original small and LR image. The SR reconstruction image obtained from the original image was enlarged to improve the image quality using triangulation interpolation until the resulting SR images were corrected. The experiment used an Intel® Core (TM) i5-4200 M 2.50 GHz processor (CPU), the Windows 10 Home Single Language 64-bit operating system, 8.0 GB of main memory (RAM), and language selection. MATLAB program development was based on the triangulation interpolation technique algorithm, nearest-neighbor method [33], SRCNN method [34], bilinear method [35], bicubic method [36], GPR method [37], and NEDI method [38]. Python was also used to test efficiency in Thai hand gesture recognition.

The experiment included three steps. First, the prototype image data were prepared. Figure 6 shows the prototype image. The image size was reduced to a LR image and re-enlarged to a large image, and image quality was improved in the SR image. The original and resulting images were then compared. The second step used triangulation interpolation to test the effectiveness of the SR image, as shown in Figure 7. The nearest-neighbor method [33], SRCNN method [34], bilinear method [35], bicubic method [36], GPR method [37], and NEDI method [38] were used as well. Image similarity errors were determined from the 1) PSNR and 2) structural similarity methods.

### 1. PSNR

PSNR, which is based on the mean square error (MSE), was used to measure image quality. This measured the effect of improving blurry images with more clarity. This value was calculated from equation 6 and equation 7 [26].

$$PSNR = 10 * \log_{10} \left[ \frac{(2^b - 1)^2}{MSE} \right] \quad (6)$$

$$MSE = \frac{1}{P_r P_c} \sum_{i=1}^{P_r} \sum_{j=1}^{P_c} (X(i, j) - X'(i, j))^2 \quad (7)$$

Where b is the number of bits in one pixel, P<sub>r</sub> and P<sub>c</sub> are the numbers of pixels in rows and columns, respectively, X(i,j) is the value of the pixel image at i,j of the original image, X'(i,j) is the value of the pixel image at i,j of the SR image.

### 2. Structural similarity

Although MSE is widely used, the results are not correlated with image quality degradation due to blurring. The structural similarity index measure (SSIM) is correlated more with visual degradation than MSE. The SSIM calculates window similarity in the overall system by including differences in the mean, variance, and correlation.

SSIM measurements were initiated with two sets of  $n$  intensities from the linked windows in the original and compared images. The mean  $\mu_x$  and  $\mu_y$  were calculated using equation 8 [26].

$$\mu_x = \frac{1}{n} \sum_{i=1}^n x_i \quad (8)$$

Where  $\mu_x$  and  $\mu_y$  are the average pixel data in window pixels  $x$  and  $y$ , respectively, and  $n$  is the intensity of two sets from the window related to the original image and comparable image.

Next, the variance  $\sigma_x$  and  $\sigma_y$  of the two sets of intensity values are calculated using equation 9.

$$\sigma_x = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \mu_x)^2} \quad (9)$$

Finally, the relationship between two sets of intensity values,  $\sigma_{xy}$ , is calculated using equation 10.

$$\sigma_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \mu_x)(y_i - \mu_y) \quad (10)$$

The mean, variance, and correlation were used to find the SSIM index with equation 11.

$$SSIM_{xy} = \frac{(2\mu_x\mu_y+c_1)(2\sigma_{xy}+c_2)}{(\mu_x^2+\mu_y^2+c_1)(\sigma_x^2+\sigma_y^2+c_2)} \quad (11)$$

where the constants  $c_1$  and  $c_2$  are not equal to zero.

The SSIM quality metrics used in this study had a window size of 8X8, and the parameters  $c_1$  and  $c_2$  were equal to 0.0001 and 0.0009, respectively. The similarity measurement in areas with all possible mean values of window offsets and all channels yielded MSSIM, which is a measure of total visual similarity. The MSSIM value was between 0 and 1, and higher values meant higher structural similarity and thus higher quality. The third step involved experiments to compare gesture recognition accuracy effectiveness in TSL.

### 3.1 Superresolution image reconstruction process

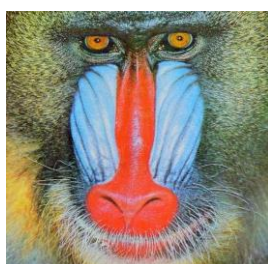
This experiment included a test set of 10 sign language images with a resolution of 512X512. The images were selected to test edge blurring, edge blocking, and more detailed visualization. Figure 9 shows the sign language and sign language 2, which researchers produced in the experiment to compare with the copyright-free signs for baboon, peppers, sailboat, airplane, bridge, Columbia, cameraman, and crowd, which came from the USC-SIPI Image Database (<http://sipi.usc.edu/database/>) [39] and have been widely used to test computer graphics performance. These experimental images contained strong or sharp edge patterns in different directions.



Sign language



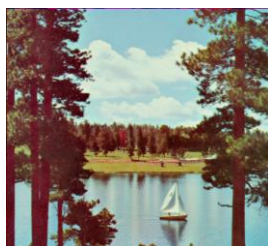
Sign Language 2



Baboon



Peppers



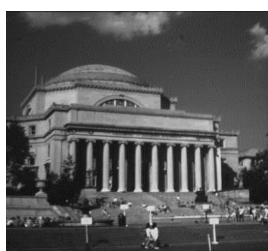
Sailboat



Airplane



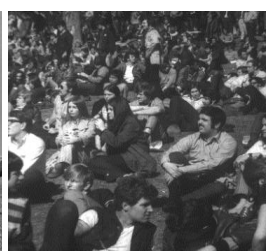
Bridge



Columbia



Cameraman



Crowd

**Fig. 6.** Original experimental images

### 3.2 Triangulation interpolation development process

This experiment developed a SR reconstruction using triangulation interpolation for feature extraction in the recognition process. The SR image reconstructed from LR images using triangulation interpolation was compared with the results from the nearest-neighbor method [33], SRCNN method [34], bilinear method [35], bicubic method [36], GPR method [37], and NEDI method [38]. The PSNR and SSIM methods were used to measure accuracy efficiency. This experiment included 15 error measurements on the selected original image and calculated the average time for all images to find the least error. The triangulation interpolation method included the following steps:

1. The original 512x512 image was acquired for the experiment. To compute the image data, the image integer data type was converted to a double format suitable for processing.
2. The original image was resized and scaled down by setting the scale value to 0.5, decreasing the image file size to 256x256. These LR images were used in the experiment using the following equations:

Width formula:

$$h' = \frac{h}{w} w' \quad (12)$$

Height formula:

$$w' = \frac{w}{h} h' \quad (13)$$

where  $h'$  is the height of ImageResize,  $h$  is the height of ImageOriginal,  $w'$  is the width of ImageResize,  $w$  is the width of ImageOriginal.

3. We created a 2D matrix to store the resized image row data equal to the row size in the resized image and a 2D matrix to store the resized image column data to define the data column equal to the resized image column size.
4. We created a 2D matrix for collecting data. The image magnification ratio was a column-size matrix (Col Pad) with a size of 256x256. The matrix stored the magnification ratio sequence (Pad) of the desired SR images to be enlarged to 512 along the column. Similarly, the extension matrix of the row (Row Pad) was 256x256 and stored the magnification ratio sequence (Pad) of the desired resolution of the maximum resolution 512 along the row.
5. We prepared a matrix for collecting new SR images. There was a new image column data matrix (New Col Unit), which stored the image size rate data and sorted the data by columns. The new image row data matrix (New Row Unit) stored the image rate and size by rows.
6. We obtained the size of the image edge to be enlarged from a LR image to a SR image to set the image edge coordinate from the lowest to the maximum edge coordinate.

7. The subimage was resized to the resized image data (ImageResize), and column expansion rate subdivision (sub Col Pad) was performed to create a column extension matrix (Col Pad). We enlarged the row ratio (sub Row Pad) to create a row extension matrix (Row Pad). Then, the new image column (sub New Col) was subdivided by creating a new column data matrix (New Col Unit) and the new row of images (sub New Row) by creating a new row data matrix (New Row Unit).
8. We constructed a matrix to collect 3D Delaunay triangulation data calculated from coordinate data points from the column extension sub (sub Col Pad) and row extension sub (sub Row Pad) by defining each point in the form of a matrix. This method connected points in space to a triangle using Delaunay triangulation.
9. Triangulation interpolation was calculated using the Delaunay triangulation matrix, sub Col Pad matrix, sub Row Pad matrix, sub Image Resize matrix, sub New Col matrix, and sub New Row matrix to create a new SR image with increased size and improved quality. The calculated data were stored in the SR image matrix (Super Image) by interpolation using triangular calculations.

Given the coordinates of the three vertices of a triangle  $P(x_1, y_1)$ ,  $Q(x_2, y_2)$ , and  $R(x_3, y_3)$ ,

$$T = (x_2 - x_1)(y_3 - y_1) - (x_3 - x_1)(y_2 - y_1) \quad (14)$$

$$M(:, 3) = ((x_1 - x_i) * (y_2 - y_i) - (x_2 - x_i) * (y_1 - y_i)) / T \quad (15)$$

$$M(:, 2) = ((x_3 - x_i) * (y_1 - y_i) - (x_1 - x_i) * (y_3 - y_i)) / T \quad (16)$$

$$M(:, 1) = ((x_2 - x_i) * (y_3 - y_i) - (x_3 - x_i) * (y_2 - y_i)) / T \quad (17)$$

Where  $T$  is the area of a triangle in the orthogonal plane,  $M$  is the matrix that collects the calculation results for each period of the points.

Thus,

$$SR = \sum_{i=1}^n O_i N_i \quad (18)$$

Where  $SR$  is a SR image that has been computed in range estimation from the sum of matrix  $O_i$ ,  $O_i$  is the original image matrix that collect the Delaunay triangle values with matrix  $N_i$ ,  $N_i$  is the matrix on which the triangle is calculated,  $n$  is the number of dimensions of the new matrix the image is scaled up to.

Figure 7 shows image reshaping to complete the SR image process. The original image was compared with the SR reconstructed image that used triangulation interpolation with the nearest-neighbor method [33], SRCNN method [34], bilinear method [35], bicubic method [36], GPR method [37], and NEDI method [38] as well as the PSNR and SSIM methods for error measurements.

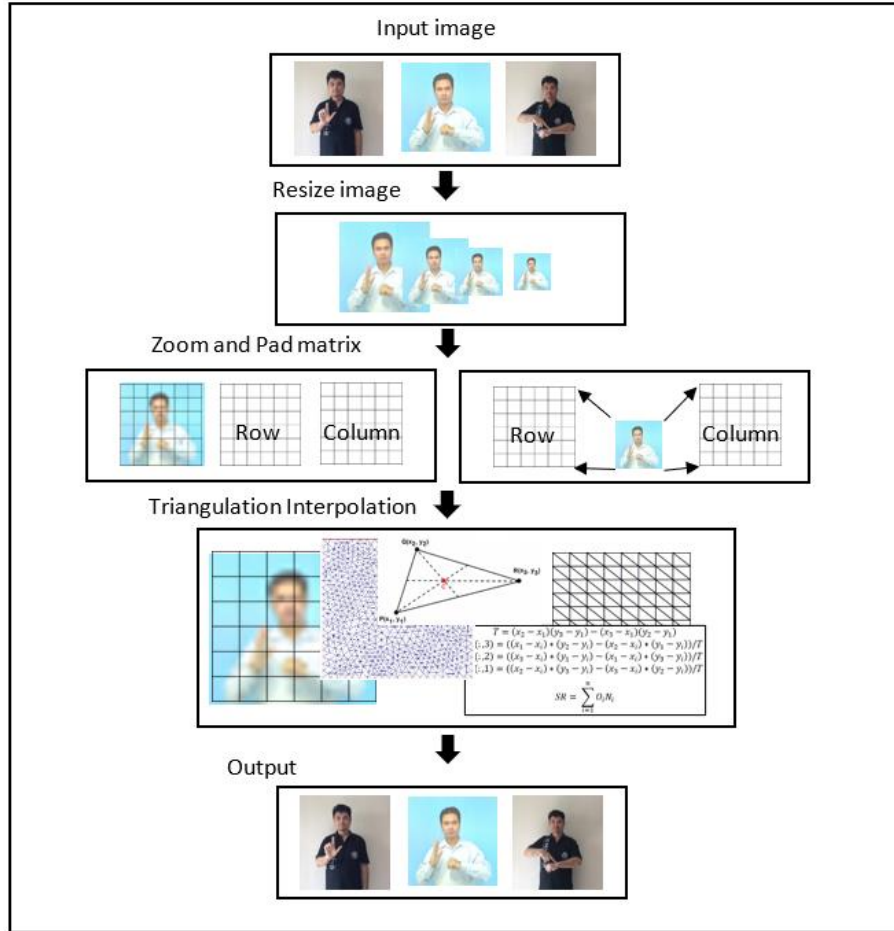


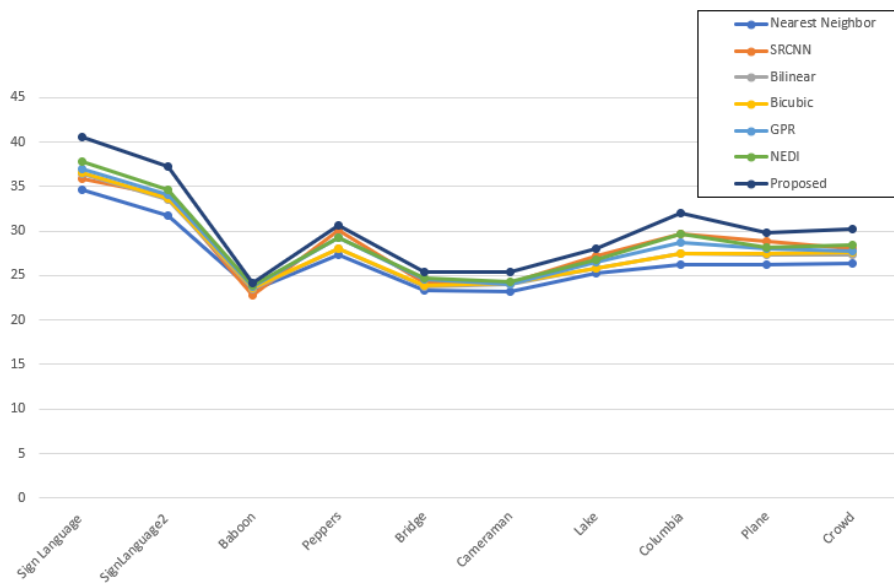
Fig. 7. Triangulation interpolation

#### 4 Experimental results and comparison

This study proposed a new approach to reconstruct SR images using triangulation interpolation for training automatic TSL recognition. The SR reconstruction images from the nearest-neighbor method, SRCNN method, bilinear method, bicubic method, GPR method, and NEDI method were compared to determine the SR image efficiency. This experiment used images of sign language and sign language 2, which the researcher produced in the experiment to compare the copyright-free signs for baboon, peppers, sailboat, airplane, bridge, Columbia, cameraman, and crowd, which came from the USC-SIPI Image Database (<http://sipi.usc.edu/database/>) [39]. We reduced these image sizes by half and used retrospective interpolation with various SR techniques to scale up images to the original size. The efficiency was measured

using PSNR and SSIM to evaluate the experimental result. The 3,000 images of three TSL gestures (“hello”, “love”, and “sick”) were used for training (1,000 images for each word), and 300 images were used for testing (100 images for each word). The images are the dataset that the researcher produced in the experiment.

These experimental results were compared using the performance of different algorithms. As shown in Figure 8 and 9, the SR algorithm efficiencies were assessed using PSNR and SSIM because each SR technique using triangulation interpolation increased image sharpness, reduced blurriness, and effectively enhanced image detail.



**Fig. 8.** SR performance comparison results using PSNR

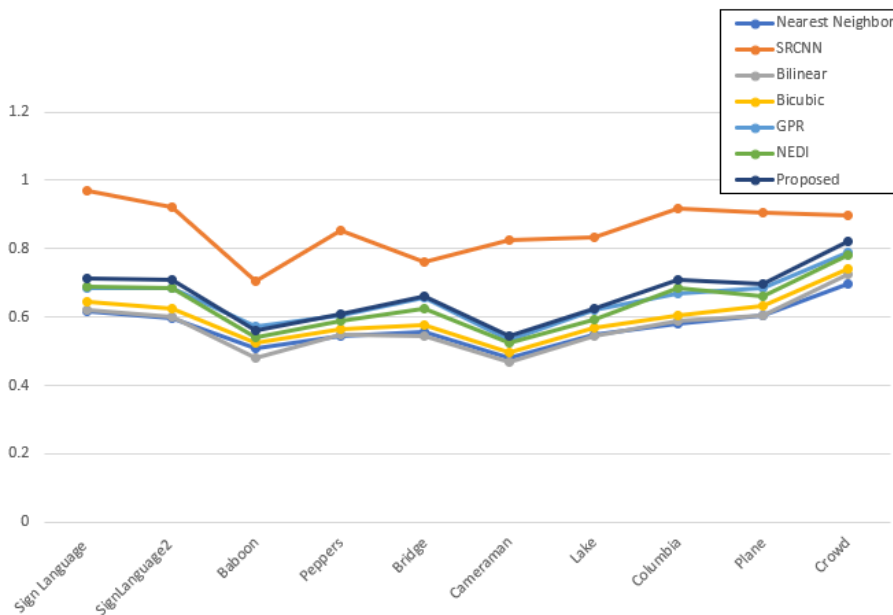
According to Figure 8, the SR performance comparison using PSNR showed that the triangulation interpolation method was better than the nearest-neighbor method [33], SRCNN method [34], bilinear method [35], bicubic method [36], GPR method [37], and NEDI method [38]. The SR process increased image sharpness, reduced blurriness, and effectively enhanced image detail.

Table 1 shows the results of the SR imaging performance test using PSNR. PSNR showed that triangulation interpolation was more effective than the nearest-neighbor method, SRCNN method, bilinear method, bicubic method, GPR method, and NEDI method. The difference between the original image and the SR image using the PSNR method for measuring image quality was 40.6081, with the sign language 2 image being 37.2952. The baboon image was 24.2063, the pepper image was 30.6391, the bridge image was 25.3810, the cameraman image was 25.3408, the sailboat image was 27.9685, the Columbia image was 31.9701, the airplane image was 29.8499, and the crowd image was 30.2059.



**Table 1.** Results of the superresolution imaging performance test using PSNR.

Images	Technique						
	<i>Nearest Neighbor</i> [33]	<i>SRCN</i> [34]	<i>Bilinear</i> [35]	<i>Bicubic</i> [36]	<i>GPR</i> [37]	<i>NEDI</i> [38]	<i>Proposed</i>
Sign Language	34.69	35.89	36.40	36.54	36.94	37.74	40.61
Sign Language 2	31.72	33.96	33.50	33.61	34.0	34.61	37.29
Baboon	23.34	22.84	23.45	23.65	23.77	23.83	24.21
Peppers	27.43	30.07	28.06	28.04	29.29	29.29	30.64
Bridge	23.35	24.19	23.81	23.87	24.53	24.66	25.38
Cameraman	23.14	24.13	23.99	24.19	24.07	24.29	25.34
Lake	25.20	27.18	25.75	25.84	26.47	26.82	27.97
Columbia	26.19	29.61	27.44	27.40	28.72	29.73	31.97
Plane	26.28	28.91	27.31	27.45	28.05	28.13	29.85
Crowd	26.39	28.03	27.36	27.54	27.6	28.37	30.21



**Fig. 9.** SR performance comparison results using SSIM

According to Figure 9, the SR performance comparison using SSIM showed that the SRCNN technique yielded the best results for reducing blurriness when scaling up images, increasing image sharpness and detail, and reducing image noise from LR image SR reconstruction. However, triangulation interpolation was one of the best options because the SRCCN method required more processing time than triangulation

interpolation as learning and feature extraction were required for the recognition process.

Table 2 shows the interpolation test results using SSIM, which calculated the SR overall similarity and difference in mean, variance, and correlation. The SSIM result showed that the SR sign language image using SRCNN for performance testing yielded the best result at 0.9686, while the sign language 2 image was 0.9232, the baboon image was 0.7040, the peppers image was 0.8540, the bridge image was 0.7595, the cameraman image was 0.8247, the sailboat image was 0.8346, the Columbia image was 0.9175, the airplane image was 0.9044, and the crowd image was 0.8960 for the efficiency measurement using SSIM. However, the triangulation interpolation technique yielded the second-best result to the SRCNN technique, as the sign language image SSIM was 0.7139, the sign language 2 image was 0.7071, the baboon image was 0.5619, the peppers image was 0.6092, the bridge image was 0.6608, the cameraman image was 0.5437, the sailboat image was 0.6227, the Columbia image was 0.7077, the airplane image was 0.6953, and the crown image was 0.8206. This result showed that the triangulation interpolation algorithm worked best with the PSNR test.

**Table 2.** Result of interpolation test using SSIM

Images	Technique						
	<i>Nearest Neighbor [33]</i>	<i>SRCN [34]</i>	<i>Bilinear [35]</i>	<i>Bicubic [36]</i>	<i>GPR [37]</i>	<i>NEDI [38]</i>	<i>Proposed</i>
Sign Language	0.62	0.97	0.62	0.65	0.69	0.69	0.71
Sign Language 2	0.59	0.92	0.59	0.63	0.68	0.68	0.71
Baboon	0.51	0.70	0.48	0.52	0.57	0.54	0.56
Peppers	0.54	0.85	0.55	0.57	0.60	0.59	0.61
Bridge	0.56	0.76	0.55	0.57	0.66	0.62	0.66
Cameraman	0.48	0.82	0.47	0.49	0.53	0.53	0.54
Lake	0.55	0.83	0.55	0.57	0.61	0.59	0.52
Columbia	0.58	0.92	0.58	0.61	0.67	0.69	0.71
Plane	0.60	0.90	0.61	0.63	0.68	0.66	0.69
Crowd	0.69	0.89	0.72	0.74	0.79	0.78	0.82



Fig. 10. Effect of SR TSL images using triangulation interpolation

## 5 Conclusions

This study offered a novel approach to determine SR performance using a triangulation interpolation algorithm for recognizing distinct sign language features. The performance was compared using sign language, sign language 2, baboon, pepper, sailboat, airplane, bridge, Columbia, cameraman, and crowd images. Our results showed that the SR technique using triangulation interpolation provided the best PSNR results. The difference between the original image and the SR image of the sign language image was 40.6084, the sign language 2 image was 37.2952, and the baboon image was 24.2063, which were the closest to the original image. For the performance measurement using SSIM, which measured structured similarity, triangulation interpolation provided the second-best SR reconstruction image result next to the SRCNN technique. However, triangulation interpolation was preferred due to its faster processing time. The TSL gesture recognition complication experiment used the 2D CNN model, and TSL gesture recognition using 2D convolution and a pooling layer were used for measuring effectiveness and accuracy. We found that the designed model increased gesture recognition with an accuracy of 0.95 and loss of 0.14 from the 2D CNNs and improved with additional training data. For the fully connected layer in the classification, although the training dataset gestures were different, the carefully selected weight in the layer preparation process greatly improved the performance.

## 6 References

- [1] Q. Zhang and J. Wu, "Image super-resolution using windowed ordinary Kriging interpolation," *Opt. Commun.*, vol. 336, pp. 140–145, 2015. <https://doi.org/10.1016/j.optcom.2014.09.060>
- [2] S. Huang, J. Sun, Y. Yang, Y. Fang, P. Lin, and Y. Que, "Robust single-image super-resolution based on adaptive edge-preserving smoothing regularization," *IEEE Trans. Image Process.*, vol. 27, no. 6, pp. 2650–2663, 2018. <https://doi.org/10.1109/TIP.2018.2809472>
- [3] R. Tsai and Thomas S. Huang, "Multiframe image restoration and registration," *Advances in Computer Vision and Image Processing.*, vol.1, no. 2, pp. 317–39, 1984.
- [4] H. Chen, X. He, L. Qing, Q. Teng, and C. Ren, "SGCRSR: Sequential gradient constrained regression for single image super-resolution," *Signal Process. Image Commun.*, vol. 66, pp. 1–18, 2018. <https://doi.org/10.1016/j.image.2018.04.012>
- [5] X. Chen and C. Qi, "Nonlinear neighbor embedding for single image super-resolution via kernel mapping," *Signal Processing*, vol. 94, no. 1, pp. 6–22, 2014. <https://doi.org/10.1016/j.sigpro.2013.06.016>
- [6] X. Li and M. T. Orchard, "New edge-directed interpolation," *IEEE Trans. Image Process.*, vol. 10, no. 10, pp. 1521–1527, 2001. <https://doi.org/10.1109/83.951537>
- [7] M. Jing and J. Wu, "Fast image interpolation using directional inverse distance weighting for real-time applications," *Opt. Commun.*, vol. 286, no. 1, pp. 111–116, 2013. <https://doi.org/10.1016/j.optcom.2012.09.011>
- [8] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, 2010. <https://doi.org/10.1109/tip.2010.2050625>
- [9] C. Dong, M. C. Leu and Z. Yin, "American Sign Language alphabet recognition using Microsoft Kinect," 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 44–52, 2015. <http://dx.doi.org/10.1109/CVPRW.2015.7301347>
- [10] R. Timofte, R. Rothe, and L. Van Gool, "Seven Ways to Improve Example-Based Single Image Super Resolution," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem, pp. 1865–1873, 2016. <https://doi.org/10.1109/CVPR.2016.206>
- [11] P. T. Hai, H. C. Thinh, B. Van Phuc, and H. H. Kha, "Automatic feature extraction for Vietnamese sign language recognition using support vector machine," *Proc. - 2018 2nd Int. Conf. Recent Adv. Signal Process. Telecommun. Comput. SIGTELCOM 2018*, vol. 2018-Janua, pp. 146–151, 2018. <https://doi.org/10.1109/SIGTELCOM.2018.8325780>
- [12] Saleh, Y., & Issa, G. F. (2020). Arabic sign language recognition through deep neural networks fine-tuning. *International Journal of Online and Biomedical Engineering*, vol. 16, no. 5, 71–83. <https://doi.org/10.3991/IJOE.V16I05.13087>
- [13] G. A. Rao, K. Syamala, P. V. V. Kishore, and A. S. C. S. Sastry, "Deep convolutional neural networks for sign language recognition," 2018 Conf. Signal Process. Commun. Eng. Syst. SPACES 2018, vol. 2018-Janua, pp. 194–197, 2018. [https://doi.org/10.1109/SPACE\\_S.2018.8316344](https://doi.org/10.1109/SPACE_S.2018.8316344)
- [14] P. V. V. Kishore, D. A. Kumar, A. Chandra Sekhara Sastry, and E. K. Kumar, "Motionlets Matching with Adaptive Kernels for 3-D Indian Sign Language Recognition," *IEEE Sens. J.*, vol. 18, no. 8, pp. 3327–3337, 2018. <https://doi.org/10.1109/JSEN.2018.2810449>
- [15] E. K. Kumar, A. S. C. S. Sastry, P. V. V. Kishore, M. T. K. Kumar, and D. A. Kumar, "Training CNNs for 3-D Sign Language Recognition with Color Texture Coded Joint Angular Displacement Maps," *IEEE Signal Process. Lett.*, vol. 25, no. 5, pp. 645–649, 2018. <https://doi.org/10.1109/LSP.2018.2817179>

- [16] M. Sadek, "A New Approach for Designing a Smart Glove for Arabic Sign Language Recognition System," 34th Natl. RADIO Sci. Conf., no. Nrsc, pp. 380–388, 2017. <https://doi.org/10.1109/nrsc.2017.7893499>
- [17] L. T. Phi, H. D. Nguyen, T. T. Q. Bui and T. T. Vu, "A glove-based gesture recognition system for Vietnamese sign language," 2015 15th International Conference on Control, Automation and Systems (ICCAS), 2015, pp. 1555-1559. <https://doi.org/10.1109/ICCAS.2015.7364604>
- [18] G. García-Bautista, F. Trujillo-Romero, and S. O. Caballero-Morales, "Mexican sign language recognition using kinect and data time warping algorithm," 2017 Int. Conf. Electron. Commun. Comput. CONIELECOMP 2017, 2017. <https://doi.org/10.1109/CONIELECOMP.2017.7891832>
- [19] Mamdouh, R., El-Khamisy, N., Amer, K., Riad, A., & El-Bakry, H. M. (2021). A New Model for Image Segmentation Based on Deep Learning. International Journal of Online and Biomedical Engineering, vol. 17, no. 7, 28–47. <https://doi.org/10.3991/ijoe.v17i07.21241>
- [20] M. D. Wibowo, I. Nurtanio and A. A. Ilham, "Indonesian sign language recognition using leap motion controller," 2017 11th International Conference on Information & Communication Technology and System (ICTS), 2017, pp. 67-72. <https://doi.org/10.1109/ICTS.2017.8265648>
- [21] J. Zhao, H. Hu, and F. Cao, "Image super-resolution via adaptive sparse representation," Knowledge-Based Syst., vol. 124, pp. 23–33, 2017. <https://doi.org/10.1016/j.knsys.2017.02.029>
- [22] A. P. Reji and T. Thomas, "Single Frame Image Super Resolution Using Learned Directionlets," Int. J. Artif. Intell. Appl., vol. 1, no. 4, pp. 29–42, 2010.
- [23] D. Glasner, S. Bagon, and M. Irani, "Super-resolution from a single image," Proc. IEEE Int. Conf. Comput. Vis., pp. 349–356, 2009. <https://doi.org/10.1109/ICCV.2009.5459271>
- [24] X. Fan, Y. Yang, C. Deng, J. Xu, and X. Gao, "Compressed multi-scale feature fusion network for single image super-resolution," Signal Processing, vol. 146, pp. 50–60, 2018. <https://doi.org/10.1016/j.sigpro.2017.12.017>
- [25] K. Chang, P. L. K. Ding, and B. Li, "Single image super-resolution using collaborative representation and non-local self-similarity," Signal Processing, vol. 149, pp. 49–61, 2018. <https://doi.org/10.1016/j.sigpro.2018.02.031>
- [26] C. Cumpim, R. Punchalard, K. Janchitrapongvej, and C. Kimpan, "Improving inverse distance weighting method for single-image super-resolution," 2016 13th Int. Conf. Electr. Eng. Comput. Telecommun. Inf. Technol. ECTI-CON 2016, 2016. <https://doi.org/10.1109/ECTICon.2016.7561489>
- [27] C. J. Enríquez-Cervantes and Ramón M. Rodríguez-Dagnino, "A Super-Resolution Image Reconstruction using Natural Neighbor Interpolation," Comp. y Sist. vol. 19, no.2, pp. 211-231, 2015. <https://doi.org/10.13053/CyS-19-2-2068>
- [28] T. Matsuo, Y. Shirai, and N. Shimada, "Automatic generation of HMM topology for sign language recognition," Proc. - Int. Conf. Pattern Recognit., pp. 3–6, 2008. <https://doi.org/10.1109/ICPR.2008.4761525>
- [29] H. G. Doan, H. Vu, and T. H. Tran, "Dynamic hand gesture recognition from cyclical hand pattern," Proc. 15th IAPR Int. Conf. Mach. Vis. Appl. MVA 2017, pp. 97–100, 2017, doi: 10.23919/MVA.2017.7986799.
- [30] C. Dong, M. C. Leu, and Z. Yin, "American Sign Language alphabet recognition using Microsoft Kinect," IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work., vol. 2015-Octob, pp. 44–52, 2015. <https://doi.org/10.1109/cvprw.2015.7301347>

- [31] X. Feng, J. Li, and Z. Hua, “Guided Filter Based Multi-scale Super-resolution Reconstruction,” *CAAI Trans. Intell. Technol.*, vol. 2, no. 2, pp. 128–140, 2020. <https://doi.org/10.1049/trit.2019.0065>
- [32] A. N. A. Rahim, S. N. Yaakob, R. Ngadiran, and M. W. Nasruddin, “An analysis of interpolation methods for super resolution images,” 2015 IEEE Student Conf. Res. Dev. SCORED 2015, pp. 72–77, 2015. <https://doi.org/10.1109/SCORED.2015.7449432>
- [33] Xu, H., Li, L., Fang, M., & Zhang, F. (2018). Movement human actions recognition based on machine learning. *International Journal of Online Engineering*, vol. 14, no. 4, 193–210. <https://doi.org/10.3991/ijoe.v14i04.8513>
- [34] C. Dong, C. C. Loy, K. He, and X. Tang, “Image Super-Resolution Using Deep Convolutional Networks,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, 2016. <https://doi.org/10.1080/21681163.2016.1154806>
- [35] X. G. Zhang, “A new kind of super-resolution reconstruction algorithm based on the ICM and the bilinear interpolation,” *Proc. - 2008 Int. Semin. Futur. Biomed. Inf. Eng. FBIE 2008*, pp. 183–186, 2008. <https://doi.org/10.1109/FBIE.2008.44>
- [36] X. Li and M. T. Orchard, “New edge-directed interpolation,” *IEEE Trans. Image Process.*, vol. 10, no. 10, pp. 1521–1527, 2001. <https://doi.org/10.1109/83.951537>
- [37] H. He and W. C. Siu, “Single image super-resolution using Gaussian process regression,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 449–456, 2011. <https://doi.org/10.1109/CVPR.2011.5995713>
- [38] J. Bae, Y. Yun, K. Kim, and J. Kim, “A new edge directed interpolation algorithm using accurate estimation of edge directional covariance,” *ISCAS 2012 - 2012 IEEE Int. Symp. Circuits Syst.*, pp. 1223–1226, 2012. <https://doi.org/10.1109/ISCAS.2012.6271456>
- [39] H. T. Wu, Z. Yang, Y. M. Cheung, L. Xu, and S. Tang, “High-Capacity Reversible Data Hiding in Encrypted Images by Bit Plane Partition and MSB Prediction,” *IEEE Access*, vol. 7, pp. 62361–62371, 2019. <https://doi.org/10.1109/ACCESS.2019.2916355>

## 7 Authors

**Eakbodin Gedkhaw** received the B.S. in computer science from Maejo University in 2004 and the M.S. in Information Technology from Naresuan University in 2009. He is currently pursuing the Ph.D. degree with the Department of Information Technology, King Mongkut’s Institute of Technology North Bangkok, Thailand. His current research interests include image processing, sign language recognition machine learning and deep learning (Email: [eakbodin.g@chandra.ac.th](mailto:eakbodin.g@chandra.ac.th)).

**Asst. Prof. Dr. Mahasak Ketcham** received the Ph.D. degree from Chulalongkorn University in 2009. He is currently an Assistant Professor with the Innovation and Technology Management Research Center, Department of Information Technology, King Mongkut’s Institute of Technology North Bangkok, Thailand. He is also a member of the ICSN2016, ICCAT’2015, ICITEE2015, ICACTCM’2014, ICMAME 2013, ICCCT2012, IMLCS2012 and ICCTAI2011. He is serving on the Editorial Boards of the *Software Networking* ISSN: 2445-9739 (Online Version) and the *Journal of Intelligent Informatics and Smart Technology*.

Article submitted 2021-11-08. Resubmitted 2021-12-19. Final acceptance 2021-12-21. Final version published as submitted by the authors.