Utilizing Mask RCNN for Monitoring Postoperative Free Flap: Circulatory Compromise Detection Based on Visible-Light and Infrared Images

CHI-EN LEE\textsuperscript{1,}, CHUNG-MING CHEN\textsuperscript{1}, (Member, IEEE), HONG-XIANG WANG\textsuperscript{1}, LI-WEI CHEN\textsuperscript{1}, AND CHERNG-KANG PERNG\textsuperscript{2}

\textsuperscript{1}Institute of Biomedical Engineering, National Taiwan University, Taipei 100, Taiwan
\textsuperscript{2}Division of Plastic and Reconstructive Surgery, Department of Surgery, Taipei Veterans General Hospital, Taipei 11217, Taiwan

Corresponding author: Chung-Ming Chen (chung@ntu.edu.tw)

This work was supported by the Ministry of Science and Technology (MOST) under Grant MOST-106-2221-E-002-006.

ABSTRACT The new postoperative free flap monitoring system combines visible-light and infrared techniques to overcome the limitations of our previous study, such as resistance to illumination change, patients with large movements, and the infrared images being too blurry to identify the boundary. In the visible-light system, the Mask region-based convolutional neural networks (RCNN) was adopted to segment the region of the free flap, and these time-course visible images were aligned using our proposed image registration method. Then, the registered visible images were projected onto the infrared image by the coordinate transformation. The analysis method adopted the residual factor analysis to extract the more unvarnished specific factors. The experiments were divided into two parts. In image processing, the accuracy of the coordinate transformation has a mean error of 1.78 pixels and a standard deviation of 0.98 pixels. The segmentation results showed excellent performance in the most severe case with apparent motion and rotation, covered by gauze and the ventilator, as well as illumination variation. The dice coefficient is 0.9551 \pm 0.0158, and the Hausdorff distance is 2.3943 \pm 0.3921 pixels. The image registration results also reveal that the Canny edge of the deformed image was superimposed onto the reference image well. In circulatory compromise detection, the vascular congestion was detected much earlier than manual observation, and the classified type of occlusion was the same as the clinical reports. Therefore, the dual-camera monitoring system provides a reliable tool for the surgeon to hold onto the chance of repairing the free flap with vascular obstruction.

INDEX TERMS Postoperative free flap, circulatory compromise, mask region-based convolutional neural networks, deep learning, image segmentation, nonrigid registration, residual factor analysis, visible-light, infrared, monitoring system.

I. INTRODUCTION

The postoperative free flap survival relies on adequate tissue perfusion. In addition to factors such as age, the severity of the injury, and whether the patient has diabetes mellitus, hypertension, and cardiac arrhythmias \cite{1}, \cite{2}, \cite{3}, \cite{4}, \cite{5}, the success rate of the free flap surgery is related to the exhibited circulatory compromise after the surgery. If a circulatory compromise cannot be detected in time, necrosis caused to the transferred flap could lead to tissue loss \cite{6}, \cite{7}, \cite{8}.
With previous studies, 5%–25% of the free flap surgeries must be remedied because of the circulatory compromise [9], [10], [11], and the probability of vein compromise was higher than artery compromise [12]. Kroll et al. [13] proposed the same result by analyzing 990 surgery cases. Chen et al. [14] analyzed 1,142 free flap surgeries, and 113 cases were remedied because of circulatory compromise. Among them, if the circulatory compromise was detected within 120 h after the surgery, the repair rate was 85%. Therefore, postoperative free flap checking is essential for decreasing the problems associated with circulatory compromise.

The evaluated rules of circulatory compromise usually examined by manual observation in the clinics are according to aspects such as its color and temperature and capillary refill, turgor, and pin-prick test results. However, the observation results are influenced by the training and experience of the medical personnel.

Moreover, the postoperative free flap observation protocol specifies that observation should be conducted once per hour for 24 h after the surgery and every 4 h for 72 h from the second day. Such observation requirements lay a heavy workload on medical personnel. Furthermore, a surgeon sometimes misses the chance to repair the free flap with vascular obstruction because some initial clinical manifestations are not obvious and thus cannot be found. Therefore, if some flap monitoring techniques are made available to the medical personnel, the observations can be more efficient and accurate. Moreover, vascular occlusion can be detected earlier.

Infrared thermal imaging devices offer advantages, such as fast operation, low cost, noninvasive behavior, contactless, and non-radiative operation. It has been considered a potential detection technique for several years. Previous studies have reported that a variation in the surface temperature of the flap and surrounding normal skin could indicate significant blood circulation problems in the flap [15], [16], [17]. In our previous animal study, 12 swine pedicle myocutaneous flaps were harvested and monitored for vascular thrombosis to demonstrate the potential of clinical monitoring using infrared (IR) cameras [18]. When the estimated error of flap surface temperature was less than 0.86°C, the sensitivity and specificity of the results were 90% and 81%, respectively. However, it was still inadequate for clinical monitoring by using IR cameras because the boundaries of the free flap in IR images were too blurred to have high quality longitudinal registration results, which would lead to inaccurate analysis results, and attaching markers on patients’ skin for a long time is impractical.

Our visible-light monitoring system with the advantages mentioned above was proposed to overcome the problem of the IR monitoring system, which conducted pixel-wise analysis and extended the application to the clinic [19]. In the clinical trial, the performance assessment of image registration reveals that the average and the standard variation of dice similarity coefficient can reach 0.959 and 0.011, respectively. The relationship between the situation of blood perfusion and the results of circulatory compromise detection at flap is summarized in Table 1, and it is shown that the detection system determined not only the moment of vascular occlusion but also the type of occlusion. However, some factors such as illumination change or the significant movement of patients would still lead to the apparent error of free flap segmentation and lower quality of image registration results. For example, the errors in identifying the free flap region and matching corresponding pairs occurred because of the patient’s large movement and illumination change shown in Figures 1(a) and 1(b), respectively. Therefore, using a visible-light system for clinical monitoring in some severe cases is also insufficient. Suppose a monitoring system with the advantages of IR and visible-light simultaneously can maintain robustness in those severe cases. In that case, the clinical application could be further extended and more valuable.

The new monitoring system in this study is expected to improve our previous systems by organizing the IR and visible-light monitoring system simultaneously and

| Red | Green | Blue | Occlusion Type |
|-----|-------|------|---------------|
| LD  | LD    | LD   | Venous        |
| LD  | LI    | LI   | Arterial      |
| SI  | SI    | SI   | Arterial      |
| SI  | LI    | LI   | Arterial      |
| SD  | SD    | SD   | Arterial      |

LI = Large Increase, LD = Large Decrease, SI = Slight Increase, SD = Slight Decrease.

*FIGURE 1. The errors of identifying the free flap region and matching corresponding pairs because of (a) the patient’s large movement and (b) illumination change.*
providing additional information for analyzing the occlusions of free flaps. However, the critical challenges for clinical application are summarized as follows:

1) The definition of the free flap boundary without markers is complex. The manual determination is labor-intensive, time-consuming, and likely to lead to interobserver differences. Moreover, the free flap segmentation method proposed in our previous study is easily affected by illumination variation and large movement.

2) Image registration would be an essential step of pixel-wise analysis because the spatial deviation between the two images needs to be eliminated, which may be generated by the patient’s large motion or rotation during monitoring. However, a free flap is a nonrigid object because of the softness and flexibility of human skin, and nonrigid registration is more complex than rigid registration because underlying nonrigid transformations are often imperceptible, complex, and challenging to model.

3) Multi-sensor data often provide complementary information for analysis. However, their point of view may be different in our system; the spatial deviation between the visible-light and IR images would lead to the failure of cooperation.

4) In a previous animal study, the variation of illumination and temperature from the environment was slight and limited, and the moment of vascular occlusion was controlled. However, those factors are variable, more complex, and unpredictable in the clinic. Thus, extracting helpful information on vascular occlusion from those factors is challenging and seriously prevents the analysis results from being influenced by environment or physiological reactions.

Therefore, the multi-sensor monitoring system aimed for clinical applications is expected to have the following abilities:

1) More robust segmentation method for the free flap: The method should identify the region of the free flap fast and precisely, even in patients with large movements or illumination variations from the environment.

2) Image coordinate transformation: The spatial deviation between the two kinds of images needs to be eliminated by establishing the relationship between them for the coordinate transformation matrix.

3) Longitudinal image registration: The sequential free flap images generated during monitoring are appropriately aligned in the spatiotemporal domain for conducting pixel-wise analysis.

4) Vascular occlusion detection: The detection system is expected to alert medical personnel and remove the influence of factors such as illumination variation, inflammation, and fever.

5) With those abilities mentioned above, the new monitoring system alleviating the effect of physiological reaction and environment would be more exact and reliable.

The analysis results may provide a great diversity of circulatory compromise detection.

In a previous visible-light study, the breakthrough of the original segmentation included extracting the features of the free flap’s color and shape from the face. Also, it is fast and accurate when patients are still and illumination is sufficient and stable. However, the alternative segmentation method is essential because of the inadequacy of clinical applications in some severe cases. In the last few years, lots of segmentation algorithms based on convolutional neural networks have been developed and applied in computer vision as well as medical image analysis [20], [21], such as region-based convolutional neural networks (RCNNs), fully convolutional neural networks (FCNs), the Fast-RCNN, and the Faster-RCNN [22], [23], [24], [25], which automatically find the solutions from the images using the combination of deep learning and computer vision, and they are evolved one after another because of expensive training and slow object detection. In 2017, the Mask RCNN was proposed and is an extended version of the Faster-RCNN model with extra mask generation and classifier networks for semantic segmentation, object localization, and object instance segmentation of images [26], which achieves pixel level segmentation and further reduces the processing time. It was the Common Objects in Context (COCO) 2016 challenge winner on single-model entry and has been widely applied to diverse image process issues. Therefore, if the Mask RCNN could supply the free flap’s faster, more efficient, and exact segmentation results in some severe cases, it may be very suitable for establishing the alternative segmentation method in this study.

Owing to the characteristic of an IR camera, the boundary of the free flap is a blur in IR images, and it limits the clinical application of IR. Thus, the concept is that the visible-light camera could be used to assist in overcoming the problem of IR images. However, the image disparity of the dual-camera system needs to be rectified before photography. The methods usually used in coordinate transformation are perspective projection and geometric transform matrix, fundamental matrix, essential matrix, and homography matrix [27]. In perspective projection, the transformation matrix parameters are difficult to estimate in real-world stereo vision because of the considerable difference between the two kinds of cameras. Moreover, if the feature points of the free flap are approximate on the same plane, then a homography matrix could implement the coordinate transformation between two cameras. For example, Jian Zhao [28] combined visible-light and thermal images by assuming the feature points of the same human as a co-plane under the pinhole camera model to establish the relationship in the homography matrix. Ji Hoon Lee et al. made visible-light and thermal cameras parallel in the horizontal direction with minimum horizontal distance. They obtained the unknown parameters of the transform matrix by the four ground truth points of the calibration plane [29]. Therefore, the homography matrix could be the appropriate method for diminishing the disparity of dual-camera systems.
To analyze the changes of the free flap in a time-course study, image registration is usually considered the essential step of aligning those images taken at different times in the spatiotemporal domain. Image registration methods are categorized into rigid registration and nonrigid registration methods. Nonrigid registration algorithms use bioimages because the human skin and tissues are soft and nonrigid [30]. However, nonrigid registration remains challenging in computer vision because of its complex nonlinear transformation models. In general, nonrigid registration includes several steps, such as feature extraction, matching, transformation, and optimization [31]. Especially in the fields of feature extraction and matching, several studies have been proposed on image registration methods with more precise results and efficient calculation.

The famous feature point detections are Moravec corner detector [32], Harris corner detector [33], and Features from accelerated segment test (Fast) corner detector [34]. In 1980, Moravec found that the difference between the adjacent pixels in a uniform image region is slight, and the difference is significantly high in all directions at the corner. In 1988, Harris improved the problem of the Moravec corner detector. It overcame the restriction of calculating only eight discrete 45° angles of the strength value. In 1998, the Fast corner detection approach was proposed and had a better performance in computation time. However, it is not as rigorous as the Harris corner detection approach, and many valuable corners are lost.

Moreover, other feature points detections, such as scale-invariant feature transform (SIFT) [35] and speeded up robust features (SURF) [36] with robustness in resisting interference of changes in scale, rotation, illumination, and local affine distortion; would also be considered. They combine a scale-invariant region detector and a descriptor based on the gradient distribution in the detected region to find the extreme points as the feature points. Furthermore, the edge of the free flap with sutures reveals the linear structure, which inspires us that the direction of sutures could be another feature. The Hessian matrix can enhance and extract the linear structure because it is based on one principal direction of the matrix [37]. To achieve the high quality of longitudinal registration results, the feature point descriptor was tried to be composed of several approaches mentioned above in this study.

In computer vision, a simple point set registration approach that directly associates the points was proposed by Scott and Longuet-Higgins [38]. However, the performance of the method is poor for nonrigid objects. Another famous point set registration method is the iterative closest point (ICP) [39]. This method iteratively calculates the least squares results of every closest corresponding point between two point sets to minimize the distance between each pair. However, ICP is unsuitable for our study because providing two point sets close to each other is impractical, especially since free flaps are nonrigid.

In contrast to ICP, coherent point drift (CPD) is a probabilistic method. In CPD, the first point set is assumed to fit Gaussian mixture models (GMMs), and the Gaussian centroids of these models are used as the initialized points of the second set [40]. Once the two point sets are optimally matched, the maximum GMM posterior probability is correspondence. Compared with other methods, CPD is more robust, especially in terms of nonlinear deformation and noise. However, patients may change their head position and thus seriously deform the flap. Moreover, some interruptions occur from the coverage area of the gauze after medical personnel changes the dressing. These issues are severe challenges to image registration.

According to the standard of observation methods, color and temperature are the principal factors in determining the situation of a free flap. For example, the color of the free flap is pale when arterial occlusion occurs and changes to dusky when veins experience congestion. Moreover, the temperature difference between the flap and control sites over a specified period is more than 1.8°C, indicating flap circulation compromise [15]. With the rules, our circulatory compromise detection method is inspired to analyze the free flap’s red green blue (RGB) intensity and temperature difference to determine when the vascular occlusion happened. However, it is easily affected by physiological and environmental factors such as illumination or temperature variation, inflammation, and fever, which may lead to false positives. Thus, extracting the critical factors from the observed data could decline the probability of a wrong warning. Factor analysis (FA) is the most popular method for data reduction to analyze a large number of correlated variables. The concept of this method is to reduce the dimensionality with a minimum loss of information and to determine potential factors with a low number of unobserved latent variables. After simplifying the description of these correlated variables, accurate results of circulatory compromise detection could be obtained.

The paper is organized as follows. Section II describes the architecture of the postoperative free flap monitoring system, including hardware and segmentation algorithms, coordinate transformation, image registration, and circulatory compromise detection. Section III summarizes the experimental studies, demonstrating that our postoperative free flap monitoring system achieves accuracy. Finally, discussions and conclusions are drawn in Sections IV and V, respectively.

II. MATERIALS AND METHODS
The dual-camera monitoring system comprises a photograph hardware module, an image processing module, and a circulatory compromise detection module. The pipeline, displayed in Figure 2, is described as follows:

1) The hardware part: Cameras would be adjusted to an initial position and aimed at the patient’s free flap before monitoring. Then, the images of the free flap were captured by a visible-light camera from different angles and transferred to a computer for training the deep learning model. After the initial setup,
visible-light and IR cameras simultaneously monitored the free flap.

2) The image processing part: First, the homography matrix of coordinate transformation between visible-light and IR cameras were calculated through calibration. Then, the free flap region of the visible-light image was segmented by the Mask RCNN and projected onto the IR image by the coordinate transformation, which substituted for the segmentation of IR with a blur boundary problem. Finally, some of the feature points extracted from the free flap were selected as corresponding points for our proposed image registration. After optimizing the registration results, the deformation images were obtained.

3) The circulatory compromise detection part: To overcome the influences of physiological or environmental factors, the registered images were analyzed using residual FA to obtain the common and specific factors. Then, the vascular occlusion of the free flap was detected by evaluating those factors’ changes.

A. HARDWARE AND EXPERIMENT SETUP

The clinical trials were verified using human flaps and approved by the Institutional Review Board of Taipei Veterans General Hospital (approval number: 2016-01-006BC). These participants were oral cancer patients who had just had free flap surgery to reconstruct defects in the head and neck region. Then, they were moved to the intensive care unit and monitored by our system for more than 24 h. During monitoring, the medical personnel would use the manual observation method for vessel occlusion detection, and they may change patients’ positions and the coverage area of the gauze.

The proposed monitoring system’s hardware comprises a visible-light camera and an IR camera. The setup of the photograph is shown in the left block of Figure 2. A visible-light camera (α6000; SONY, Tokyo, Japan; 6000 × 4000 pixels) and an IR camera (spectrum 9000 MB; United Integrated Services Company, Taiwan; 320 × 240 pixels, spectral range 7–14 µm, sensitivity 0.05°C∼0.08°C) were operated simultaneously, and the arm can be manually adjusted by medical personnel.

The experiments are divided into the image process part and compromise detection part, which are described as follows.

1) EVALUATION OF IMAGE PROCESSING

a: COORDINATION TRANSFORMATION
A mask regarded as the prosthesis of the face was used to estimate the homography matrix between the two kinds of cameras, which were placed closely on the arm and kept a fixed distance from the mask. Several markers were attached to the mask to establish the transformation matrix, and the volunteers implemented the evaluation experiment.

b: THE SEGMENTATION BY MASK RCNN
Approximately 50 images of the patient’s free flap as the training data were captured from different angles to simulate the patient’s movement. The segmentation results were evaluated with influences such as patents’ movement, illumination variation, and gauze coverings and ventilators.

c: LONGITUDINAL IMAGE REGISTRATION
This is to demonstrate the validity of employing the visible-light camera to assist the image registration of IR images. Through matching the corresponding pairs and
deformation as well as coordination transformation, the image registration results of IR images evaluated the spatial deviation between two time-course images.

2) THE DEMONSTRATION OF CIRCULATORY COMPROMISE DETECTION
The moment of vascular occlusion is the most critical information in compromise detection. Using the proposed dual-camera system, the detection of vascular occlusion would be demonstrated and compared with medical personnel’s check. Moreover, the analysis results may be influenced by physiological or environmental factors during the monitoring. Thus, the elimination of those factors by residual FA also needs to be demonstrated.

B. COORDINATE TRANSFORMATION
There are two assumptions in the dual-camera system to simplify the complexity of coordinate transformation. First, both images view the same plane from a different angle. Second, the camera is rotated about its projection center without any translation. Thus, the coordinates of two images are coincident by camera calibration based on the homography matrix. If the corresponding points $X$’ and $X$ between two images are matched, the linear mapping is described by the homography matrix as follows:

$$X' = HX$$  \hspace{1cm} (1)

where $X' = (x', y', 1)$, $X = (x, y, 1)$, and $H$ is a $3 \times 3$ matrix with eight degrees of freedom known as homography matrix. Then, the equation is rewritten as follows:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$  \hspace{1cm} (2)

$$x' = \frac{h_{11}x + h_{12}y + h_{13}}{h_{31}x + h_{32}y + h_{33}}$$  \hspace{1cm} (3)

$$y' = \frac{h_{21}x + h_{22}y + h_{23}}{h_{31}x + h_{32}y + h_{33}}$$  \hspace{1cm} (4)

If setting $h_{33} = 1$, the homography matrix with eight variables needs at least four corresponding sets to be solved, which are represented as follows:

$$x' = h_{11}x + h_{12}y + h_{13} - h_{31}xx' - h_{32}xy'$$  \hspace{1cm} (5)

$$y' = h_{21}x + h_{22}y + h_{23} - h_{31}yy' - h_{32}xy'$$  \hspace{1cm} (6)

where $(x_1', y_1', 1)(x_1, y_1, 1)$, $(x_2', y_2', 1)(x_2, y_2, 1), \ldots (x_n', y_n', 1)(x_n, y_n, 1)$ are the selected four corresponding sets between the two images to obtain the homography matrix.

C. IMAGE SEGMENTATION BY MASK R-CNN
The Mask R-CNN framework model is shown in Figure 3, which is divided into two stages: the first stage, formed in the left green block, inputs the images into the backbone network model (deep residual network, ResNet101 [41] + feature pyramid network (FPN) [42]) to extract features and generate the corresponding feature maps; the second stage is that the region of interest (ROI) is obtained by the region proposal network (RPN) [25], and then the flow goes through the ROI Align to generate the fixed size feature map. Eventually, the flow is separated into two branches on the right side of Figure 3. One branch enters the fully connected layer for object classification and frame regression, and the other one enters the entire convolution network (FCN) for pixel segmentation [23]. The workflow is described in detail as follows.

1) TRAINING METHOD
In the training process, the patient was monitored using our system for 24 h and captured one image per minute to build the dataset. Then, 50 images selected as the training data from the dataset presented the patient’s movement during monitoring, as shown in Figure 4. The remaining data were used to be the training data and validation data. Moreover, the training data rotated at different angles ($10^\circ$, $20^\circ$, $35^\circ$, and $340^\circ$) to simulate the possible angles of photography. After that, the network model’s weights were initialized and obtained from pretraining on the COCO [43] dataset, which reduced the number of images needed to train the network and decreased the time required to train the models. Eventually, the training and validation data inputted the Mask R-CNN for training. The segmentation results of testing data are evaluated in Section III.

2) BACKBONE NETWORK
The backbone network for the Mask R-CNN is a multilayer neural network to extract high-level visual features from the entire image. It usually adopts several network structures such as ResNet, Visual Geometry Group [44], and dense convolutional network [45], and ResNet was chosen as the backbone network. ResNet is typically divided into ResNet50 and ResNet101 with depths 50 and 101, respectively, which rely on the size and complexity of the object for adoption. In this study, the ResNet101 was considered the backbone network and expected to work well. Moreover, the mechanism of the backbone network can employ the pretrained parameters and does not need a training network from random initial parameters, which transfers the trained COCO dataset through transfer learning [46].

3) FPN
In deep neural networks, the lower layers’ features pass through many network layers to reach the top layers. After passing through many layers, some of the lower-level information may be lost. However, the information in the
lower-level features is essential, for instance, segmentation. Thus, the FPN as a top-down pyramid architecture extracts high-level features from the first pyramid, which are presented as the five layers \{C1, C2, C3, C4, C5\} and pass to lower layers using lateral connections to obtain the feature maps of four feature levels \{P2, P3, P4, P5\} shown in the left of Figure 3.

4) REGION PROPOSAL NETWORK
The RPN was proposed in Faster-RCNN to replace the selective search method by predicting proposals from each sliding window location in the feature map generated by the backbone network. These proposals are parameterized relative to the bounding boxes (named anchors), centered at the sliding window, and associated with different sizes and aspect ratios. Then, the anchors for candidate boxes are selected by comparing the Intersection over Union between different anchors and the ground truth. The highest score of candidate boxes with the highest probability of being from the foreground class is chosen. Moreover, the nonmaximum suppression is used to screen out the optimal anchor as the preselected bounding box. Eventually, the final proposal area (ROI) is obtained by employing the bounding box regression method to fine-tune the bounding box.

5) ROI ALIGN AND NETWORK HEAD
To prevent the misalignment caused by ROI pooling in Faster-RCNN, the ROI Align was proposed to adopt the bilinear interpolation to compute the exact values of the input features at four regularly sampled locations in each ROI bin. Then, it performs max or average pooling on the features.

Finally, the ROI Aligned features are passed to the network head, which performs three parallel tasks of bounding box regression, classification, and masks prediction.
D. LONGITUDINAL IMAGE REGISTRATION

In this study, image registration was implemented in the visible-light images and generated the deformation model, also used in the IR image registration. The flowchart of the longitudinal image registration algorithm is shown in Figure 5. The image registration system performs four processes: feature point extraction (comprising Hessian matrix and SIFT), matching (CPD), affine transformation [47], and gradient vector flow (GVF) snake deformation [48].

1) FEATURE POINT EXTRACTION

a: THE FEATURE POINTS FROM THE MORPHOLOGICAL INFORMATION

The skeleton structure of the edge of the free flap was generated by evaluating the eigenvalues of the Hessian matrix, which can provide morphological information about the sutures around the free flap. The crosspoints and endpoints of the skeleton were identified as the feature points. The skeleton structure of the edge of the free flap was generated by evaluating the eigenvalues of the Hessian matrix,

\[
H = \begin{bmatrix}
\frac{\partial^2 I(x,y)}{\partial x^2} & \frac{\partial^2 I(x,y)}{\partial x \partial y} \\
\frac{\partial^2 I(x,y)}{\partial y \partial x} & \frac{\partial^2 I(x,y)}{\partial y^2}
\end{bmatrix},
\]

where \( I(x,y) \) is the original IR image, and VR is used to quantify the response to the characteristic values of the Hessian matrix as follows:

\[
VR = \exp \left( \frac{\lambda_1}{s_2^2} \right)^2 \left( 1 - \exp \left( \frac{\lambda_2^2}{2s_2^2} \right) \right),
\]

where \( \lambda_1 \) and \( \lambda_2 \) represent the smallest characteristic absolute value and the largest characteristic absolute value, respectively. Subsequently, the threshold of the VR index must be determined to obtain the image represented by \( \lambda_1 \) and \( \lambda_2 \). Eventually, MATLAB was used to identify the middle lines of the skeleton by applying the command “bwmorph,” and the crosspoints and endpoints of the skeleton can be easily detected by considering pixels with more than three neighbors and only one neighbor, respectively.

b: THE FEATURE POINTS OF SIFT

The feature points of SIFT were detected in two steps: extracting and filtering. In the first step, the image was convolved using Gaussian filters at different scales, and the differences between these consecutive Gaussian-blurred images were then determined. The difference in Gaussian (DoG) of the image between different scales, \( k \sigma \) and \( k_2 \sigma \), is as follows:

\[
D(x,y,\sigma) = L(x,y,k_1 \sigma) - L(x,y,k_2 \sigma),
\]

\[
L(x,y,k \sigma) = G(x,y,k \sigma) * I(x,y),
\]

where \( G(x,y,k \sigma) \) is the Gaussian function at scale \( k \sigma \) and \( I(x,y) \) is the original image. Once the DoG images have been obtained, feature points are identified as the local minimum or maximum of the DoG images within different scales. However, too many feature points were produced, and some were unreliable. Thus, the second step involves filtering out points that exhibited a low contrast or were poorly localized along an edge. Eventually, the feature points of the free flap are shown in Figure 6. The red points are the feature points of SIFT, and the green points are the crosspoints and endpoints of the Hessian matrix. Several feature points of the edge are extracted. In addition, the texture of the skin is also extracted as the feature point.

2) MATCHING

After extracting the feature points, the corresponding pairs were determined through CPD. The template point set \( Y = (y_1, \ldots, y_M)^T \) (expressed as a \( M \times 2 \) matrix) can be aligned with the reference point set \( X = (x_1, \ldots, x_N)^T \) (expressed as
The template point set $Y$ is supposed to be the centroid of the Gaussian mixture model with the coherence constraint, and the reference point set $X$ serves as the data point that must be fitted. If the transfer function $X = T(Y, \theta)$ exists, then the energy function can be written as follows:

$$E(Y) = -\sum_{n=1}^{N} \log \sum_{m=1}^{M} e^{-\frac{1}{2} \left\| \frac{x_n - y_m}{\sigma} \right\|^2} + \frac{\lambda}{2} \phi(Y)$$  \hspace{1cm} (12)$$

where $\lambda$ is a weighting constant and $\phi(Y)$ is a smooth function of motion. The aim is to minimize the energy function through estimation maximization to determine the parameters $\theta$ and $\sigma$. Then, the corresponding pairs between the two point sets can be found.

3) AFFINE TRANSFORMATION
To reduce the complexity of the computation, a linear transformation model was employed to establish the deformation model in this study. Affine transformation comprised rotations, translations, and dilations, and shears is a linear map method that is easy to implement and time-saving. Affine transformation can be represented as follows:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} s_x \cos(a) & -s_y \sin(a) \\ s_y \sin(a) & s_x \cos(a) \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix}$$ \hspace{1cm} (13)$$

where $(x', y')$ and $(x, y)$ are the corresponding pairs between the 1st and nth IR images; $s_x$ and $s_y$ are the scale factors along the $x$ and $y$ axes, respectively; $sh_x$ and $sh_y$ are the shear factors along the $x$ and $y$ axes, respectively; $a$ is the angle of rotation; and $\Delta x$ and $\Delta y$ are the displacement along the $x$ and $y$ axes, respectively. After these parameters are solved, the deformation model is obtained.

4) GVF SNAKE DEFORMATION
A snake model was used to optimize the deformation results to make the reference one’s deformed free flap edge fit [49]. However, the snake model had poor convergence to the boundary concavities. Thus, the GVF snake model was used to overcome the limitations and replace the constraint energy of the traditional snake model [48]. The GVF snake equation is represented as follows:

$$\alpha x''(s) - \beta x^{m''}(s) + w(x, y) = 0$$ \hspace{1cm} (14)$$

where $\alpha$ and $\beta$ are the weighting parameters that control the snake’s tension and rigidity, respectively, and $x''(s)$ and $x^{m''}(s)$ denote the second and fourth derivatives, respectively, of $x(s)$ with respect to $s$, which is the curve’s function.
FIGURE 9. The segmentation results in a free flap. (a) The boundary of the free flap was segmented well, even if the patient had an apparent motion and rotation. (b) In contrast to a previous study, the boundary of the free flap segmented by Mask RCNN was more precise and robust against the influence of covering.
The abovementioned equation is defined as the GVF field, and the energy function should be minimized.

\[ \varepsilon = \int \int \mu (u_x^2 + u_y^2 + v_x^2 + v_y^2) + |\nabla f|^2 |w - \nabla f|^2 \, dxdy \]  

(15)

where \( \mu \) is a regularization parameter; \( u_x, u_y, v_x, \) and \( v_y \) are derivatives of the vector field along the x and y axes; and \( f(x, y) \) is an edge map derived from the original image. Therefore, the second term dominates the integrand when \( |\nabla f| \) is large. The second term was minimized by setting \( w = \nabla f \).

**III. EXPERIMENTAL RESULTS**

**A. ERROR EVALUATION FOR COORDINATION TRANSFORMATION**

The process of camera calibration is shown in Figure 8. The mask with several markers were used to estimate the homography matrix between the two images in Figures 8(a) and 8(b). The corresponding points with blue numbers selected from the markers were employed to solve the matrix. After the homography matrix was obtained, the accuracy was evaluated by the experiments of the participants. In Figure 8(c), the region covered by the gauze was formed by four points with a green number, which was projected onto the IR image of Figure 8(d) through the estimated homography matrix. The projected points, red points, were used to evaluate the deviation between the two images. The mean error was 1.78 pixels, the standard deviation was 0.98 pixels, and the unit was 1.8 mm per pixel. Even though the standard deviation is a little high because of the difference between the prosthetic and the real face, it is still reasonable for coordination transformation.

**B. THE SEGMENTATION RESULTS OF MASK RCNN**

A patient underwent a free flap surgery and was monitored by our system for 24 h. During monitoring, 1440 visible-light images were captured for testing, 250 images were captured from different angles and used as training data, and 100 images were used as validation data. The segmentation
FIGURE 11. The residual factor analysis (FA) results of a patient. (a) A patient with venous congestion and the influence of physiological and environmental factors. The free flap temperature explicitly reveals the moment of vascular occlusion at time 140 through residual FA. The color of the free flap was dusky at time 170 (red dotted line), which indicated venous congestion. Venous congestion occurred considerably earlier than the time noted by the nurse (red arrow). (b) Patient without vascular occlusion. The specific factor of the red green blue (RGB) data was flat and stable.
FIGURE 12. The segmentation results in the free flap. The boundary of the free flap was segmented well, even if the patient had an apparent motion, rotation, and illumination variation.

results are shown in Figure 9(a). The boundary of the free flap was well segmented, even if the patient had an apparent motion and rotation. Even when a part of the free flap was covered by gauze and the ventilator, the segmentation results were still exact. In contrast to that in a previous study, it was easily affected by the patient’s movement and covering. The comparison is presented in Figure 9(b). The right figure represents the unsuccessful segmentation case of a previous study. The boundary of the free flap was identified ambiguously and did not match the proper position.

In contrast to that in the previous study, the boundary of free flap segmented using the Mask RCNN is more precise and robust against the influence of covering. The dice coefficient is 0.9551 ± 0.0158, and the Hausdorff distance is 2.3943 ± 0.3921 pixels, as summarized in Table 2.

C. THE IMAGE REGISTRATION RESULTS IN IR IMAGES

After obtaining deformed images in the visible-light of our image registration system, they were projected onto the IR image by the coordinate transformation, which is shown in

| TABLE 2. The evaluation of the segmentation results. |
|-----------------------------------------------|
|               | Mean | Max  | Min  | Standard deviation |
| Dice coefficient | 0.9551 | 0.9832 | 0.8869 | 0.0158             |
| Hausdorff Distance | 2.3943 | 4.6804 | 1.4142 | 0.3921             |
Figure 10. The Canny edge of the source image was superimposed onto the reference images, and the slight spatial deviation revealed the errors from image registration and coordinate transformation, which is still reliable for image registration. Therefore, the visible-light camera successfully assisted the IR image registration to overcome the blur boundary problem.

D. THE CIRCULATORY COMPROMISE DETECTION IN THE CLINICAL TRIAL

In the clinical trial, a patient underwent free flap surgery and was monitored using our system from July 12, 2019. The doctor and nurse verified that one of the two veins in the flap was congested at 14:09 on the same day (red arrow), and then the patient had surgery to repair the free flap with vascular obstruction. The residual FA results were selected from 10:22 to 14:12 on July 12. As shown in Figure 11(a), the IR and visible-light data revealed the free flap’s temperature and color (decomposed into the RGB data). The common physiological and environmental influence factors were successfully removed from the original data. In particular, the specific temperature factor explicitly revealed the moment of vascular occlusion at time 140 through residual FA. Although temperature rises at time 180 because of some complicated physiological reactions, the warning of obstruction was much earlier than visible-light and medical personnel’s check. Moreover, the specific factor of RGB also indicated that the color of the free flap was dusky at time 170 (red dotted line). The identification of occlusion type in Table 1 indicated venous congestion and was the same as the clinical reports.

To verify the reliability of the detection system, the patient without vascular occlusion was monitored using our system. The residual FA results were selected from 2:59 to 04:36 on January 12, 2018, as displayed in Figure 11(b). The specific factors of temperature (blue) and color (red) were found to be flat and stable. This implies that vascular occlusion did not occur. Therefore, our detection system is accurate and robust.

IV. DISCUSSION

A. THE EVALUATION OF COORDINATION TRANSFORMATION IN AN IDEAL CASE

To improve the coordination transformation, the individual homography matrix of participants rather than a mask could minimize the error of coordination transformation. In this simulated experiment, the individual homography matrix was directly established by attaching several markers to the volunteer’s face. Then the error of the estimated homography matrix was calculated through the deviation between the projected and original markers in the IR image. The mean error was 1.05 pixels, and the standard deviation was 0.45 pixels, which indicated that the error of the individual homography matrix was almost half of the estimated homography matrix by a mask. However, it is impractical and difficult to keep markers attached to the face of patients during monitoring. Therefore, we aim to improve the coordination transformation model in the coming future.

B. THE SEGMENTATION RESULTS WITH ILLUMINATION VARIATION

In a previous study, the illumination variation was the primary environmental factor influencing the visible-light system, leading to the failure to segment the free flap’s boundary. Thus, the segmentation results of the Mask RCNN were evaluated without eliminating the illumination change. As shown in Figure 12, the boundary of the tree flap was segmented well, even if the patient had an apparent motion, rotation, and illumination variation. The segmentation method could resist the influence of covering part of the free flap with gauze and a ventilator. In contrast to the previous study, the results revealed that the segmentation method is more practical, precise, and robust.

V. CONCLUSION

The two postoperative free flap monitoring techniques, including the visible-light and IR cameras, have several advantages such as fast operation, low cost, noninvasive behavior, contactless, and non-radiative operation. It is difficult to find all the benefits mentioned above in most detection techniques in the clinic. However, factors such as illumination change or patients with large movements lead to an obvious error of free flap segmentation and lower image registration quality in the visible-light system. Moreover, the boundaries of the free flaps in IR images were blurry to have a high quality of longitudinal registration, which would lead to inaccurate analysis of the results, and attaching markers on patients’ skin for a long time is impractical. Thus, the new postoperative free flap monitoring system that combines one visible-light camera and one IR camera has been designed to overcome the abovementioned problems.

The Mask RCNN was adopted in the visible-light system to segment the free flap region. These time-course visible images were aligned using our proposed image registration method for conducting pixel-wise analysis to prevent the manual determination of the boundary of the free flap. Then, the registered visible images were projected onto the IR image by the coordinate transformation, which was substituted for the blur boundary problem in IR images. In circulatory compromise detection, the residual FA was used to eliminate the common factors, such as the influence of physiological or environmental factors. Specific factors were left unvarnished for analysis.

The experiments were divided into two parts: the evaluation of image processing and the demonstration of circulatory compromise detection. In image processing, a mask with several markers was used to estimate the homography matrix between the visible-light image and the IR images. The accuracy was evaluated by the participants, and the mean error was 1.78 pixels, and the standard deviation was 0.98 pixels. Then, the boundary of the free flap was segmented well by the Mask RCNN. Even if the patient had
an apparent motion and rotation, part of the free flap was covered by the gauze, and the ventilator and the influence of illumination variation were included, the segmentation results were still exact. The dice coefficient is $0.9551 \pm 0.0158$, and the Hausdorff distance is $2.3943 \pm 0.3921$ pixels. After that, our image registration system obtained the deformed visible-light images and projected them onto the IR image. The evaluation revealed that the Canny edge of the source image was superimposed onto the reference images well. In circulatory compromise detection, vascular congestion was detected much earlier than manual observation. In particular, the IR system’s warning was even earlier than that of the visible-light system. Moreover, the classified type of occlusion was the same as in clinical reports. Therefore, the dual-camera monitoring system overcome the limitations of the previous studies, extended the clinical application, and provided a reliable tool for relieving the workload on medical personnel and the surgeon to hold onto the chance of repairing the free flap with vascular obstruction.

REFERENCES

[1] M. B. Hirigoyen, M. L. Urken, and H. Weinberg, “Free flap monitoring: A review of current practice,” *Microsurgery*, vol. 16, no. 11, pp. 723–726, 1995.

[2] J. A. Gussenoff, S. J. Vega, S. Jiang, A. B. Behnam, H. Shitany, H. R. Herrera, A. Smith, and J. M. Serletti, “Free tissue transfer: Comparison of outcomes between university hospitals and community hospitals,” *Plastic Reconstructive Surg.*, vol. 118, no. 3, pp. 671–675, Sep. 2006.

[3] K.-D. Wolff, F. Hölzle, A. Wysluch, T. Mücke, and M. Kesting, “Incidence and time of intraoperative vascular complications in head and neck microsurgery,” *Microsurgery*, vol. 28, no. 3, pp. 143–146, 2008.

[4] A. A. Chalil, T. D. Anderson, G. S. Weinstein, and R. S. Weber, “Internal jugular vein versus external jugular vein anastomosis: Implications for successful free tissue transfer,” *Head Neck*, vol. 23, no. 6, pp. 475–478, Jun. 2001.

[5] D. A. Hidalgo, J. J. Disa, P. G. Cordeiro, and Q.-Y. Hu, “A review of 716 consecutive free flaps for oncologic surgical defects: Refinement in donor-site selection and technique,” *Plastic Reconstructive Surg.*, vol. 102, no. 3, p. 722, 1998.

[6] J. S. Brown, J. C. Devine, P. Magennis, P. Sillifant, S. N. Rogers, and E. D. Vaughan, “Factors that influence the outcome of salvage in free tissue transfer,” *Brit. J. Oral Maxillofacial Surg.*, vol. 41, no. 1, pp. 16–20, Feb. 2003.

[7] D. T. Bui, P. G. Cordeiro, Q.-Y. Hu, J. J. Disa, A. Pusic, and B. J. Mehrara, “Free flap reevaluation: Indications, treatment, and outcomes in 1193 free flaps,” *Plastic Reconstructive Surg.*, vol. 119, no. 7, pp. 2092–2100, Jun. 2007.

[8] V. Punchapakases, P. Addison, E. Beausang, J. E. Lipa, R. W. Gilbert, and P. C. Neligan, “Role of thrombolysis in free-flap salvage,” *J. Reconstruct Microsurg.*, vol. 19, no. 8, pp. 523–530, 2003.

[9] T. Kubo, K. Yano, and K. Hosokawa, “Management of flaps with compromised venous outflow in head and neck microsurgical reconstruction,” *Microsurgery*, vol. 22, no. 8, pp. 391–395, 2002.

[10] D. A. Hidalgo and C. S. Jones, “The role of emergent exploration in free-tissue transfer: A review of 150 consecutive cases,” *Plastic Reconstructive Surg.*, vol. 86, no. 3, p. 492, 1990.

[11] M. A. Schusterman, M. J. Miller, G. P. Reece, S. S. Kroll, M. Marchi, and H. Goepfert, “A single center’s experience with 308 free flaps for repair of head and neck cancer defects,” *Plastic Reconstructive Surg.*, vol. 93, no. 3, pp. 772–794, 1994.

[12] D. Novakovics, R. S. Patel, D. P. Goldstein, and P. J. Galliano, “Salvage of failed free flaps used in head and neck reconstruction,” *Head Neck Oncol.*, vol. 1, no. 1, p. 33, Dec. 2009.

[13] S. S. Kroll, M. A. Schusterman, G. P. Reece, M. J. Miller, G. R. D. Evans, G. L. Robb, and B. J. Baldwin, “Timing of pedicle thrombosis and flap loss after free-tissue transfer,” *Plastic Reconstructive Surg.*, vol. 98, no. 7, pp. 1230–1233, Dec. 1996.
[39] P. J. Besl and D. N. McKay, “Method for registration of 3-D shapes,” Proc. SPIE, vol. 1611, pp. 586–606, Apr. 1992.

[40] A. Myronenko, X. Song, and M. Carreira-Perpinan, “Non-rigid point set registration: Coherent point drift,” in Proc. Adv. Neural Inf. Process. Syst., 2007, pp. 1–8.

[41] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 770–778.

[42] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, “Feature pyramid networks for object detection,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 2117–2125.

[43] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollar, and C. L. Zitnick, “Microsoft COCO: Common objects in context,” in Proc. Eur. Conf. Comput. Vis. Cham, Switzerland: Springer, 2014, pp. 740–755.

[44] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” 2014, arXiv:1409.1556.

[45] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 4700–4708.

[46] L. Torrey and J. Shavlik, “Transfer learning,” in Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques. Hershey, PA, USA: IGI Global, 2010, pp. 242–264.

[47] Y. Lamdan, J. T. Schwartz, and H. J. Wolfson, “Affine invariant model-based object recognition,” IEEE Trans. Robot. Autom., vol. 6, no. 5, pp. 578–589, Oct. 1990.

[48] C. Xu and J. L. Prince, “Gradient vector flow: A new external force for snakes,” in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., Jun. 1997, pp. 66–71.

[49] M. Kass, A. Witkin, and D. Terzopoulos, “Snakes: Active contour models,” Int. J. Comput. Vis., vol. 1, no. 4, pp. 321–331, Jan. 1987.

**CHI-EN LEE** received the B.S. degree in electrical engineering from the National United University, Miaoli, Taiwan, in 2007, and the M.S. degree in communication engineering from the National Chin Nan University, Nantou, Taiwan, in 2009. He is currently pursuing the Ph.D. degree with the Institute of Biomedical Engineering, National Taiwan University, Taiwan. His research interests include image registration, pattern recognition, and signal process.

**CHUNG-MING CHEN** (Member, IEEE) received the Ph.D. degree in electrical engineering from Cornell University, Ithaca, NY, USA, in 1993. He subsequently joined the Center for Biomedical Engineering, National Taiwan University, Taiwan, where he worked as a Research Assistant Professor. He is currently a Professor with the Department of Biomedical Engineering, National Taiwan University. He has published more than 113 journal articles and 160 conference papers. His research interests include medical image analysis, machine learning/deep learning approaches to assisting in early detection, differential diagnosis, prognosis prediction of diseases, and IR medical imaging. He is currently an Associate Editor of the Biomedical Engineering: Applications, Basis and Communications journal. He had served as a program committee member or an organizing committee member for several international conferences.

**HONG-XIANG WANG** received the B.S. degree in biomedical engineering from Chung Yuan Christian University, Taoyuan, Taiwan, in 2018, and the M.S. degree in biomedical engineering from the National Taiwan University, Taiwan, in 2020. His research interests include image registration, image segmentation, and automatic control.

**LI-WEI CHEN** received the B.S. degree in biomedical imaging and radiological science from China Medical University, Taichung, Taiwan, in 2015, and the M.S. degree in biomedical engineering from the National Taiwan University, Taiwan, in 2017, where he is currently pursuing the Ph.D. degree with the Institute of Biomedical Engineering. His research interests include image segmentation, pattern recognition, and machine learning.

**CHERNG-KANG PERNG** received the B.S. degree from the School of Medicine, National Yang Ming Chiao Tung University, and the Ph.D. degree from the Department of Biomedical Engineering, National Yang Ming Chiao Tung University. He is currently the Division Chief of Plastic and Reconstruction Surgery at Taipei Veterans General Hospital. He is also an Assistant Professor with the Department of Medicine, National Yang Ming Chiao Tung University. His expertise is head and neck reconstruction, breast reconstruction, lymphedema reconstruction, limb reconstruction, microsurgery, cosmetic surgery, and laser treatment filler injection.

VOLUME 10, 2022