Supervised Scene Illumination Control in Stereo Arthroscopes for Robot Assisted Minimally Invasive Surgery

Shahnewaz Ali, Yaqub Jonmohamadi, Yu Takeda, Jonathan Roberts, Ross Crawford, Ajay K. Pandey

[1]Robotics and Autonomous Systems, School of Electrical Engineering and Robotics, Queensland University of Technology, Gardens Point, Brisbane, QLD 4001, AUSTRALIA.

***Abstract*— Minimally invasive surgery (MIS) offers many advantages to patients but it also imposes limitations on surgeons ability as no tactile or haptic feedback is available. From medical robotics perspective, visualizations issues specific to MIS such as limited field of view and the lack of automatic exposure control of the surgical area make it challenging when it comes to tracking tissue, tools and camera pose as well as in perceiving depth. Lighting plays an important role in 3D reconstruction and variations due to internal illumination conditions are known to degrade vital visual information. In this work, we describe a supervised adaptive light control system to solve some of the image visualization problems of MIS. Our proposed method is able to classify underexposed and over-exposed frames and adjust lighting condition automatically to enrich image quality. Our proposed method uses support vector machines to classify different illumination conditions. Visual feedback is provided by gradient information to assess image quality and justify classifier decision. The output of this system has been tested against two cadaver knee experiment data with an overall accuracy of 97.75% for under-exposed and 89.11% for over-exposed classes. Hardware implementation of this classifier is expected to result in adaptive lighting for robot assisted surgery as well as in providing support to surgeons by freeing them from manual adjustments to lighting controls.**

***Index Terms*— Intelligent Light Intensity Control, Support Vector Machine, Knee Arthroscopy, MIS, 3D Reconstruction, Robotic-Assisted Surgery.**

I. Introduction

Contrary to open surgery, minimally invasive surgery (MIS) scenes are observed through a miniaturized endoscopic camera inserted into the body through a small incision and only a small region of the surgery scene becomes visible. Moreover, this procedure provides no haptic and tactile sensing hence, vision is the only source of information available to surgeons. In other words, surgeons have to operate on 3D surface based on 2D images projected on screen. This combined with a limited field of view adds further complexity to MIS as a surgical procedure. In recent years, many studies have aimed at increasing the accuracy for MIS [1-5]. It is usually well established that vision system plays an important role and robust vision is essential to avoid unintentional injury during MIS. Some recent studies suggest that simply extending the visualization of surgical area is able to significantly increase surgery accuracy and reducing operating time and 3D reconstruction of the surgical scene is desired to further improve surgical outcomes [1,6,7]. Furthermore, 3D reconstruction is expected to have a great impact on the robotic-assisted surgery, such as tissue - tool tracking, and pose estimation.

Stereoscopic vision, simultaneous localization and mapping (SLAM) and photometric stereo are widely used to reconstruct 3D surfaces along with the alternative methods such scanning technology [42,43,44,45,48]. All these methodologies often encounter difficulties primarily due to the complex optical nature of anatomic constructs and also due to the lack over the lighting control [1,48,49]. It is also true for other computer vision algorithms such as segmentation [50]. Additionally, different MIS process encounters different level of constraints. As an example, arthroscopy is limited to dimension constraints due to the lack of bone-joint space where laparoscopic MIS process can have a bit larger dimension. A small change to light intensity can lead to an image becoming over- or under-exposed and it is widely encountered problem in arthroscopy. Saturated image parts contain almost zero information, no features, and in situations like this stereo correspondence search converges

This work was conducted as a part of the Australia-India Strategic Research Fund (Project AISRF53820)

Shahnewaz Ali is with the Queensland University of Technology, 2 George St, Brisbane, QLD 4000, Australia (e-mail: shahnewaz.ali@hdr.qut.edu.au).

DrYaqub Jonmohamadi is with the Queensland University of Technology, 2 George St, Brisbane, QLD 4000, Australia (e-mail: y.jonmo@qut.edu.au).

Dr Yu Takeda is with the Queensland University of Technology, 2 George St, Brisbane, QLD 4000, Australia (e-mail: yu.takeda@qut.edu.au).

Prof. Jonathan Roberts is with the Queensland University of Technology, 2 George St, Brisbane, QLD 4000, Australia. (e-mail: jonathan.roberts@qut.edu.au).

Prof. Ross Crawford is with the Queensland University of Technology, 2 George St, Brisbane, QLD 4000, Australia (e-mail: r.crawford@qut.edu.au).

Ajay Pandey is with the Queensland University of Technology, 2 George St, Brisbane, QLD 4000, Australia (e-mail:a2.pandey@qut.edu.au).

to an ambiguous result that lead to fail the whole reconstruction process. Partial reconstruction can be a possible solution when light intensity is changed to a lower value. Similar problem also arises when a scene contains low light intensity that is often marked as an underexposed or dark image. Apart from the 3D reconstruction process, other vision related approach such as segmentation and registration, localization and pose estimation also encounter same problems that causes the whole system to fail [1].

In conclusion, the light source or lighting condition has a great influence on the MIS vision enhancement procedures thus on the accuracy of the MIS procedure as a whole. Surgeons or robotic platforms require to have a stable vision of the internal surgery scene. Considering knee arthroscopy, the presence of the diffuser (water), the anatomical structure, and camera motion are the most common factors that restrict to have a constant light intensity. Moreover, the control system must respond as fast as possible due to the camera motion. To have a robust surgical vision this fundamental problem has to be solved.

II. Problem Formulation

During a MIS procedure, often the vision system and tissues has to remain in a close proximity. Considering anatomical structure, in some MIS procedures such as in knee arthroscopy, tissues further away from the arthroscope often receive less light that causes formation of dark pixels (under exposed) whilst near objects encounter pixel saturation (over exposed). This situation is depicted in Fig.1. Although the overexposed and the underexposed events frequently occur in all MIS procedures, in current work this problem is analyzed in the context of the knee arthroscopy. The anatomical structures present inside the knee cavity are generally characterized to be consist of either bone or soft tissues.

Considering curved bone structure and narrow spaces, different areas of a surgical scene could receive different levels of illumination; which result in over and under exposures. Example of this scenario is shown in Fig.2. Moreover, some tissues reflect more light than others. A representative example is the Anterior Cruciate Ligament (ACL) that often causes pixel saturation. In addition, the distal parts of the anatomy, e.g. bony joint structure, create a more complex situation as depicted in Fig. 1-(a) where some parts are shown to become over exposed and under exposed in the same frame. Generally, parts that are closer to the camera or arthroscope tip reflects more light. This leads to an interesting scene visualization problem. While the total scene intensity may remain the same, increasing the light intensity saturates the closer pixels and decreasing light increases underexposed problem for the distant parts. Automatic methods to balance both over exposed and under exposed fractions with minimum frame loss are therefore much desired. These types of scene or anatomical regions of the knee model requires balanced lighting condition that minimizes total amount of underexposed and overexposed region or requires an adjusted lighting condition based on the priority of that region according to the clinical needs. Hence, it is essential to ignore the parts having less priority. Additionally, for 3D reconstruction using stereo cameras, we have noticed that changing light intensity above what is generally considered a normally exposed scene can further improve the depth estimated by stereo arthroscope.

Lighting control plays a critical role in MIS visualization. Light intensity is controlled manually and surgeons have to adjust light intensity multiple times that interrupts with the surgery process flow. Conventional endoscopic camera systems mostly provide monocular vision. Monocular image saturation is a relatively simple problem than those encountered in stereo image pairs and stringent controls are needed for correct depth estimation. Clearly, manual approach to lighting control is not feasible for robotic-assisted surgery. Robotic imaging systems would require robust lighting control for depth estimation.

In this work, we present a supervised light control system that imitates like an endoscopic camera illumination system and provides real time feedback to improve stereo system accuracy and robustness by recommendations to adjust scene illumination condition to enhance scene features and image contrast even if the scene is considered at normal exposure. We have trained and validate our system with two distinct cadaver experiments.

Fig. 1: Stereo arthroscope frames showing (a) tissue interfaces between femur and meniscus. (b) mostly soft tissue such as ACL. Saturated and dark pixel contain little to no color and texture information. Images are taken from the video sequences of cadaver knee experiment. Point A represents saturated pixels caused by the soft tissue (near object). Point B represents underexposed part caused by the bone cavity or due to occlusion caused by tissue arrangement (far object) and point C represents shadow parts (underexposed).

III. Related Work

At present, no work has been found so far that resolves the illumination problem automatically in arthroscopic or even endoscopic imaging systems. A big volume of articles has been

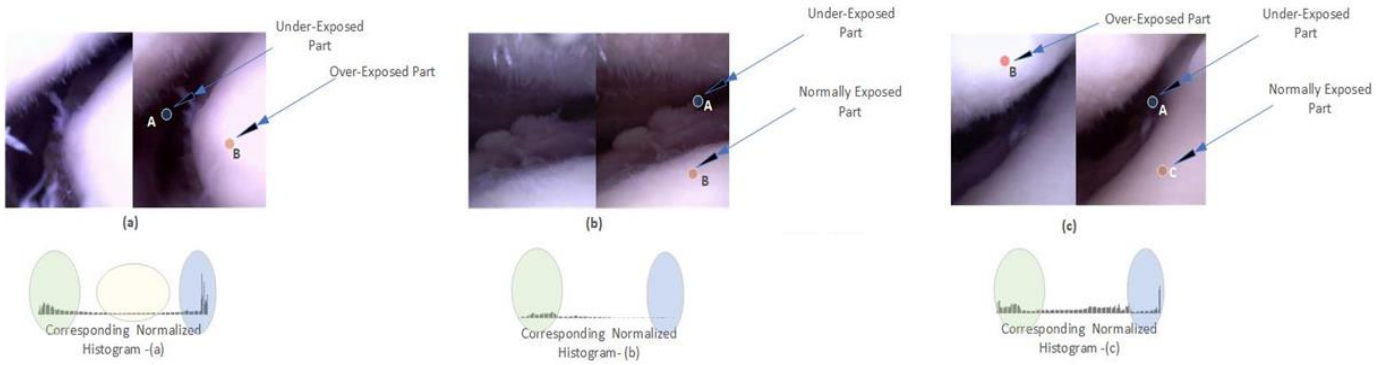


Fig. 2: Representation of different anatomical scenes of the knee cavity viewed through the stereo arthroscope and corresponding normalized histogram for each stereo pair. (a) depicts images where the bone surfaces are close to the imaging system and became overexposed where the internal anatomy inside the cavity remains invisible. The corresponding histogram has high distributions in both overexposed and underexposed end while less so on the middle tone. In the clinical perspective priority is given to the bone surfaces. The bone cavity scene (b) represents the similar phenomena but in this situation the anatomy inside the cavity is slightly visible (point A) whereas bone surface (point B) is about to be saturated. The corresponding histogram has high distribution on the underexposed region and little to no distribution on the saturated region where middle tone shows apparently few distributions. A little increase in light intensity will turn the bone surface overexposed which is not recommended for safe navigation of the tool. In this type of the scene, the optimal decision is not to change the light intensity but to navigate the tool inside the cavity in order to make the obscure surgical scene visible. Scene (c) also depicts the similar situation. According to the histogram, it has high distribution on the underexposed regions but in order to navigate the tool inside the cavity, the priority is given to the bone surface in order to steer the endoscope inside the bone space. In this proposed method, the classifier learns these characteristics of the knee model indirectly through the statistical distributions during the training phase to decide expected illumination condition.

found to control camera exposure time by controlling the amount of light that falls on an image sensor. It has some similarities related to MIS visualization problems stated above. However, it is worth noting that most of these approaches are based on static images. In this context, camera response versus scene light intensity is evaluated through the equation described below.

$$\sum = K * L * G * S \qquad (1)$$

Where, $\sum$ is the sum value of imaging signal, K is a constant, L is the brightness of the object, G is the gain of the automatic gain control (AGC) circuit, S represents shutter speed of the camera. The automatic exposure bracketing based method is widely used to adjust exposure time [10,12,13,14]. Pourreza et.al presented an automatic exposure selection method based on exposure bracketing [10]. It requires n number of image frames with different exposure values. In their method, they used two building blocks namely scene analysis and exposure time. The clustering method was used to divide the image frame into three parts and exposure time was calculated through the camera response function. Radiometric based exposure control has been proposed by Kim and et.al. for outdoor scene analysis [11]. The radiometric camera response function is individual to the camera sensor [15,16] that often generates camera specific solution. However, this method also requires a set of still images. A statistical measure such as mean, mean sample value, etc. are used to intercept scene intensity distribution for normal exposure images. Impact of lighting on image formulation is significant, some literatures address light noise conditions with the recommendation to the use of filters, artificial intelligence algorithms and different level of optoelectronics devices [46,47].

Alternatively, neural networks have been studied to estimate chromaticity of image intensity [20]. Geoffroy et.al. proposed a method that uses convolutional neural network (CNN) to estimate illumination intensity for High Dynamic Range images. In their work CNN was used to predict illumination for low dynamic range images. Usually a CNN network requires a large number of data sets for training. Additionally, deep learning requires a dedicated number of processing units to achieve real time performance [21].

Some other methods assess image quality at different exposure values [22]. Images having different exposure values can be achieved by transforming an image using gamma correction or tone mapping curve [22,23]. In MIS context these approaches are not always valid. Camera motion, tissue fiber movement, non-uniform illumination, surface curvature and distance between camera and surface limits the utility of almost all static image-based approaches. In endoscopic imaging process we need to minimize the frame loss, whereas exposure bracketing methods always have a number of redundant frames and camera navigation process has to remain in a fix position for multiple times for image stabilization. Moreover, most of the literature does not consider underwater or turbid environment encountered in MIS. In this context, scattering is a one of the most considerable problem where light falls back to the image sensor even without any interaction with the object of interest [17]. Same illumination level respect to in air, in underwater environment can cause blur or saturation due to scattering. Considering the above context, the rapid movement of the camera through the surgery process could lead visualization system deficiencies.

In short, structure of internal anatomy plays a crucial role and it takes the whole problem to different scale from the visualization approaches that are basically focused on usual indoor or outdoor environments. However, scene intensity conveys a pivotal information regarding overexposed and underexposed situation, whereas in MIS domain we have noticed that at same intensity we can have both situations present in a frame. These circumstances limit the utility of scene intensity-based approaches. However, we have considered these approaches as background literature to validate our

concept.

IV. Methodology

We defined the illumination control problem as a classification problem here. Therefore, the aim of this work is to identify whether the classification and regression methods can infer requirements specific to the level of illumination desired in visualizing the surgical scene in terms of light distribution. The whole scene light conditions are grouped into three classes as follows; I) overexposed, II) underexposed, and iii) normal. In many surgery scenes, these first two scene categories are overlapped.

This method provides a solution that establishes robust illumination condition for the stereo and the monocular endoscope. We define 'area of interest' as an overlapped region between the stereo image pair captured by the stereo endoscope developed in our lab. Area of interest plays a significant role on stereo matching process which is the leading mechanism of stereo vision. It is more reasonable that the area of interest preserves sufficient image context.

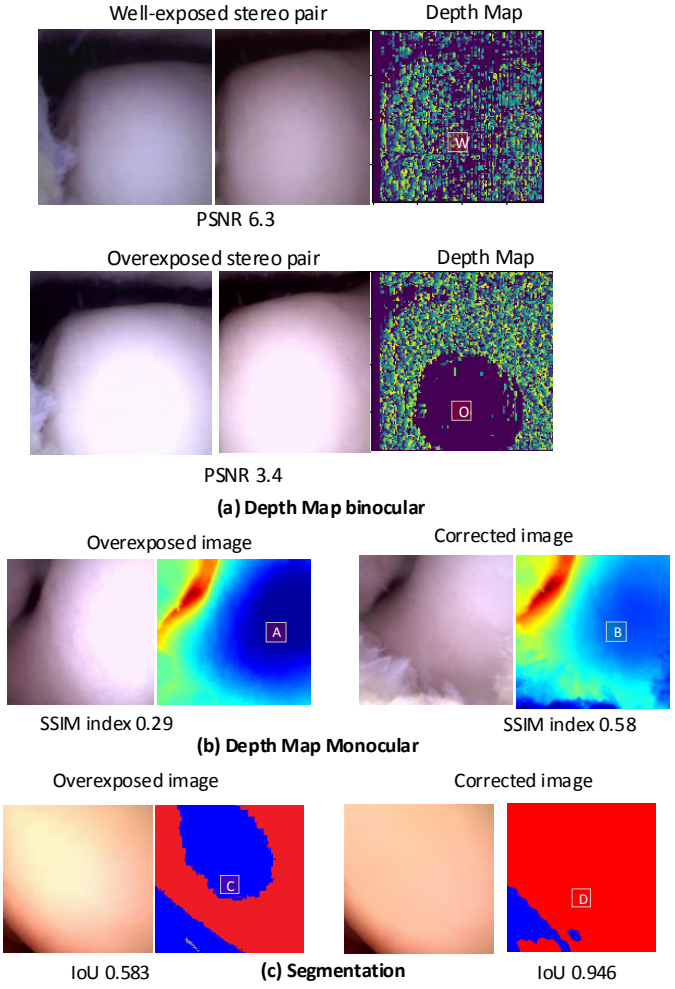


Fig. 3: Representative images that express the effect of saturated pixels on the corresponding depth map in binocular and monocular vision. (a,b). Area marked with point A,O are overexposed region as a consequence's depth discontinuity is observed. Hence, Structural Similarity index (SSIM) and Peak noise to ration (PSNR) shows poor result where the right image (b) does not experience overexposed pixels and its SSIM is 0.58. (c) Similarly, segmentation process shows poor outcomes with overexposed area described by point C where with well exposed image denoted by point D on right image shows better accuracy. Accuracy metric intersection over union (IoU) for overexposed image is 0.58 where well exposed achieved 0.94.

Image pixel values are grouped into three groups and different threshold values were defined to serve this purpose. Usually threshold values are set by experiments [38-40]. In this work, eight-bit gray scale threshold values are defined empirically through the observations of intensity responses of internal anatomy using vision processing algorithm such as rank transformation and stereo matching followed by the human visibility limits. Three thresholds are defined as follows;

$$I(x,y) = \begin{cases} Overexposed, if\ 255 \geq I(x,y) > 226 \\ Underexposed, if\ \ 101 \geq I(x,y) \geq 0 \\ NormallyExposed, if\ 226 \geq I(x,y) > 101 \end{cases} \quad (2)$$

Here, $I(x,y)$ represents pixel intensity of the image coordinate x, y. The maximum intensity of a saturated pixel in eight-bit gray scale level is 255. However, when a pixel is about to be saturated it has the same effect on the human visual system or on the vision processing methods as it is depicted by Fig. 3. The underexposed pixels also have the same effect. Research found that in middle gray tone pixels receive optimal exposure value which is 128 in eight-bit gray scale level [41]. In order to identify optimal value of these two extremes, arthroscopy video sequences are segmented in to two slices where applicable. Slices are marked as an overexposed and underexposed. A range of gray level value then applied on the whole video sequences in order to find the optimal gray level intensity values for these two extremes. When a certain gray level intensity achieved a maximum segmentation accuracy, then that gray level value is selected as a threshold value.

In this work, intersection over union (IoU) is used to evaluate segmentation accuracy at different threshold values. Underexposed, overexposed and normal indices describe the overall image statistical information. Underexposed index expresses relative underexposed pixel amount over the whole image.

Similarly, overexposed and normal exposed indices express relative amount of over and normally exposed values over the whole image. These three indices define the whole dataset in 2D space.

Support Vector Machine (SVM) is used with radial basis function (gaussian kernel). SVM was originally proposed by V. N. Vapnik [8]. SVM is widely used to address classification as well as regression problem as it is described in literature [51]. It is a binary classifier but in conjunction with other SVM it can perform multi classification problem. It classifies data with possible highest inter-class distance. Unlike multi-layers perception such as deep learning networks, SVM can be trained with limited dataset. It consumes fewer hardware resources compared to deep network, yet it provides efficient classification result in our problem domain. Instead of implementing multi-layer perception and conditional probability computation, SVM provides support vectors that are used to compute class distance from the hyperplane in run time. Moreover, Light control system has to operate in real time with minimal frame lost. Trained SVM has these strong points.

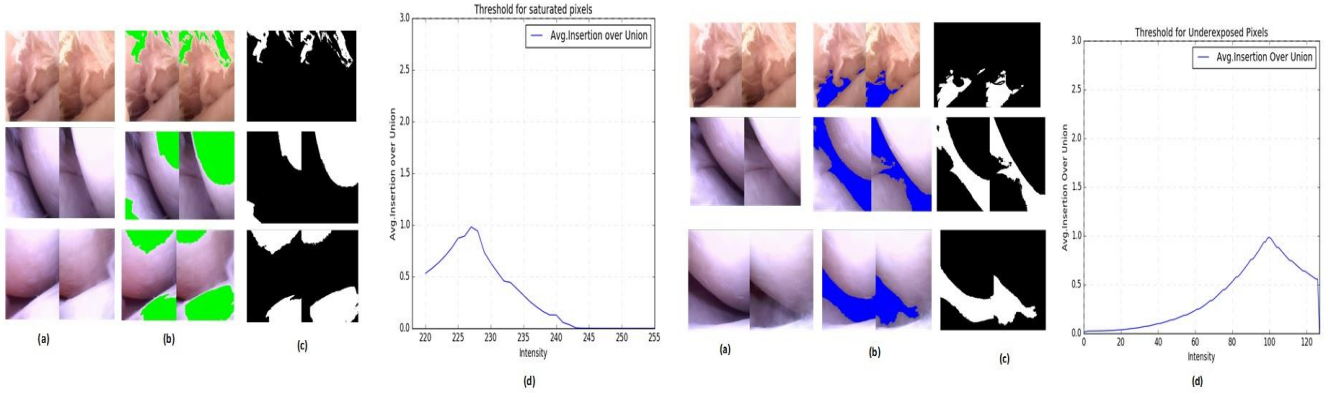
The most complex problem is to identify overexposed and underexposed scenes. To classify these total three lighting

condition (underexposed, normal, overexposed), SVM with radial basis function (RBF) is applied that is stated in equation (3).

$$K(x, y) = e^{-\gamma|x-y|^2} \quad (3)$$

The advantage of the radial basis function is that for given sufficient data, it is able to define an efficient hyperplane discriminant among the nonlinear data set. It does not have saturation problem. The nature of lighting distributions is non-linear as depicted in Figure 6, we chose nonlinear kernel function for the proposed classifier. Comparative study shows radial basis function outperforms with nonlinear data among others such as polynomial kernel [52,53]. Random subset from shuffled sample dataset is compared against different non-linear kernel before RBF kernel is selected for this approach. Mean accuracies of cross validation dataset are 0.9559448, 0.82038217, and 0.98152866 for the SVM kernel Sigmoid, Polynomial and Radial basis function respectively. Optimization takes places to establish classification and marginal error tolerance depicted in Figure 6.

Feedback system can justify classifier outcomes in real environment. In a sufficient light conditions image contains more context features, such as edges. Based on this, feedback system evaluates present image context and past image context. That can be achieved by evaluating image features such as edges, corner. In underwater environment, it has been observed that two consequent image frames show a subtle gradient change. Therefore, an empirical threshold is required that calibrates the system. Figure 4 presents our approach to define individual threshold values across all image conditions. For each frame different level of gray scale intensity values are used to achieve binary segmentation. The outcomes of each threshold value are then compared to the ground truth mask and



(I) **(II)**

Fig. 4: Set **(I)** represents overexposed threshold selection process. Left most images (a) represents the original images captured by the stereo endoscope, the next column (b) represents the segmented images and the column (c) represents corresponding mask for each stereo image pair. Over exposed pixels are marked with green color on the segmented images and those pixels are manually selected for each video frames. (d) represents average IoU curve and the optimal value is obtained at intensity level ~226 value is selected. Likewise, Set **(II)** represents underexposed threshold selection process. Left most images (a) represent the original images captured by the stereo endoscope, the next column (b) represents the segmented images and the column (c) represents corresponding mask for each stereo image pair. Underexposed pixels are marked with blue color on the segmented images and those pixels are manually nominated for each video frames. (d) represents average IoU curve and the optimal value is obtained when intensity level ~101 is selected.

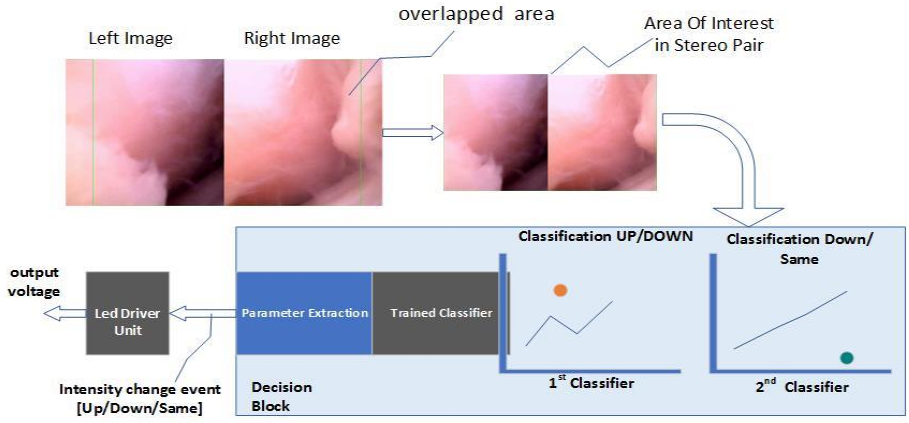


Fig.5: System overview of light control system. Stereo pair is cropped into areas of interest. In the next step, overexposed index, underexposed index and normally exposed index are calculated that are used by the SVM classifiers.

then IoU (Jaccard index) metrices are calculated. In the next step, overexposed index, underexposed index and normally exposed index is calculated that are used by the SVM classifiers, as detailed below in Figure 5. 1st SVM classifier determines whether scene need more illumination or less. If scene requires more illumination second SVM classifier determines whether it actually needs more illumination or maintain the current illumination condition. Figure 6. shows hyperplane and optimized SVM (RBF) margin for over and under exposed conditions and output to identify light intensity changes between two classes.

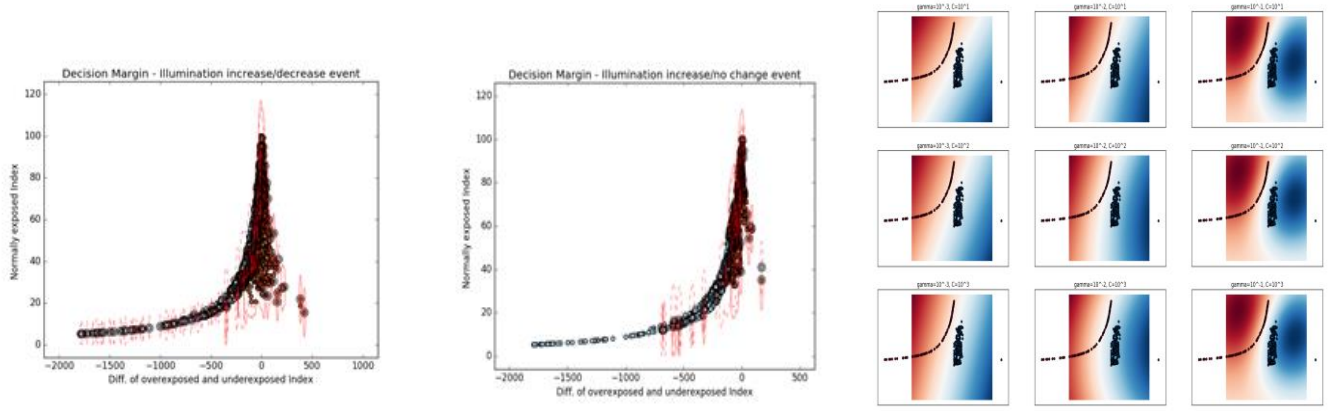


Fig. 6: Hyperplane and optimized SVM (RBF) margin with subset of our sample data to identify light intensity change. **(a)** left image represents classifier margin of two events; intensity increase and intensity decrease. **(b)** middle represents classifier margin of two events; intensity increase and intensity no change classes.(**c**.) right image expresses the example optimization process of SVM parameter, Gamma and C. Different values of gamma and c are used to select optimal values of them in order to make the class separable while avoiding marginal error.

In this work, any scene context changes are referred as a motion. Pixel wise comparison is futile due to illumination variation, displacement and rotations. Additionally, frame similarity index such as structural similarity index (SSIM) may not provide a good result in this situation. Moreover, most of these approaches are based on gray level image statistics such as mean and variance. To compare two pixels under different illuminations, in this article census transformation in grayscale is used.

V. Training

For training purpose, two cadaver knee experiment video data are used. Each frame is manually labeled. Scene is marked as an overexposed image if it contains lack of image features due to pixel saturation otherwise it is label as an underexposed image. If the scene is classified as an overexposed image, it is then further marked as a normally exposed image if it contains enough features and contains minimum image saturation otherwise overexposed image due to dark pixel amount. Additionally, if image does not contain saturated pixels and increasing light intensity enhances image quality than it is labeled as an underexposed image even though it may not contain enough dark pixels. This training data set is used to train the second SVM.

Further SVM regularization is recommended to increase classifier accuracy. Initially approx. 2950 frames are used to create training dataset. Optimization process is performed over RBF gamma and C parameter to fit optimal classification boundary curvature and the penalty value for misclassification. The optimal gamma value is 0.01 and optimal C value is 1 for the classifier up (light intensity increase) and no change (light intensity no change) classes. Similarly, for the classifier of the intensity change up (increase) and down (decrease), the method shows optimal performance on gamma equal to 0.001 and C equal to 100. The training accuracy are approx. 97.75% and 89.11%. Higher gamma value is discarded to avoid data overfitting problem. During the optimization process minimization of classification error is more considered rather than the marginal error.

A. Data Collection

During our cadaver experiment, stereo video sequences of knee Arthroscopy are collected. Stereo endoscope has manual light control unit. During this arthroscopy surgeon changes light intensity several times in different part of the knee cavities. Those video sequences are used to validate model. Moreover, one set of dedicated video sequences has been recorded that contains a linear light intensity changes from low to high in some position during the camera trajectory.

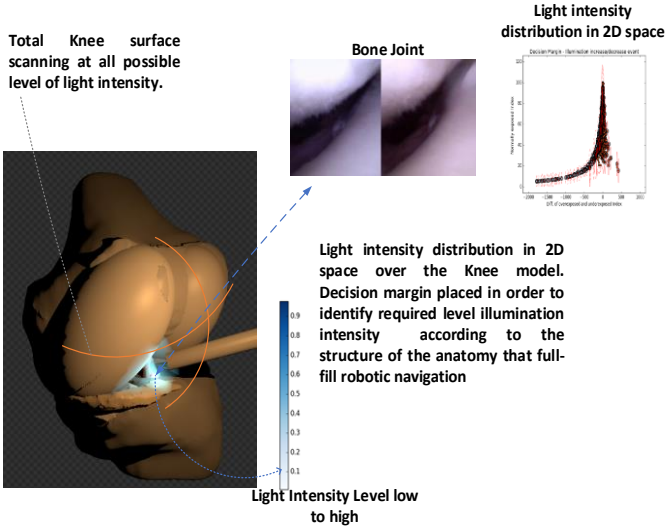


Fig. 7. Stereo endoscope along with its illumination sources are used to scan the whole knee anatomy. The surface light distributions and the ratio of it to the different classes at all possible intensity levels are extracted to construct a 2D data set. This data set is used to train our classifier.

VI. Design of the Endoscopic Stereo Camera

Figure 7 presents the clinical context and a representative location of arthroscope within the knee cavity during MIS. Details of this stereo camera prototype developed in our lab is shown in Figure 8. A commercially available traditional arthroscope currently used by surgeons is also shown in this

figure. Our stereo camera prototype is comprised of following elements;

- a pair of muC103A cameras together with their C8262 UVC interface modules,
- two white LED (T0402W) for illumination,
- 3D printed camera head for mounting cameras and the LED,
- 3D printed box for containing the wires and the circuits, and
- insertion tube.

Fig.8. The stereo endoscope prototype (a) together with the close-up view at the tip (b) and the 3D design (c) and the muC103A camera (d). The potentiometer shown by the red arrow is used to adjust the LED intensity. The endoscope circuits and extra wiring are contained inside the black box and connected to the computer using two USB cables.

muC103A is a CMOS technology camera sensor from OmniVision which can stream video at 30 fps with 400*400-pixel resolution and the field of view in 120 deg. The tip diameter of 1.52 mm makes this camera ideal for the endoscopic applications. The two cameras were mounted on the 3D printed head manually, hence some degree of misalignment between the two cameras was unavoidable. The T0402W LED consumes less than 20 mA at 3.3 V and is 1 mm wide. The baseline for the cameras was set 1.52 mm (distance between the optical center of two cameras is 1.52 mm). By considering the 87.5-degree field of view of the muC103A, the amount of overlap between the stereo pair is approximately 79% on 10 mm away from the stereo cameras.

TABLE I
CLASSIFIER TRAINING AND VALIDATION

Number of Training Frames and Dataset *Cad.	Training Accuracy	Number of validations Frames
Frames: 250 File id: 2019-02-26-12-04-54	98.5	60 frames Above 98 % accuracy
Frames: 300 File id: 2019-02-26-12-03-38	99.62	100 frames
Frames: 600 File id: 2018-09-14-11-33-23	99.78	200 frames
Frames: 900 File id: 2018-09-14-11-19-47	98.24	300 frames
Frames: 1500 File id: 2018-09-14-11-43-57	98.87	500 frames

Total 3,350 frames are used for training and 1,160 frames are used for validation.
*Internal Dataset, Knee Arthroscopy Experiment (Cadaver)

VII. RESULT

Total 3,350 video frames are selected from our cadaver knee experiments. Distorted and contaminated frames are discarded. During this training phase we select ⅓ of frames from each video file that contains different levels of difficulties such as shadow, bone joint and cavity, close contact between the camera and tissue etc. We use those frames for cross validation purposes. Table I shows the all outcomes during our training phase. During the test phase, fully trained classifier is tested against different arthroscopy video data along with cross validation data set. In some video frames different stereo camera is used. All the SVM graphs are created using scikit [30].

This methodology is compared against the camera exposure control methods in order find a benchmark. Camera exposure control system limits the amount of light that falls on digital camera sensor in order to achieve good quality of images. In this evaluation, increase of exposure time simulated as an increase of light intensity, decrease of exposure time as a decrease of light intensity, and no effect of exposure time interpreted as a no change of light intensity. Compared methods are mainly based on histogram analysis and scene quality analysis. However, compared to natural images MIS images has a number of limitations such as camera locomotion, distance between the camera surface, and no uniform light distribution etc.

Algorithm 1 is based on Gradient Information [22,35]. Gradient information is computed as follows;

$$\bar{m} = \begin{cases} \frac{1}{n}\log(\lambda(m_i - \delta) + 1)\, for\ m_i \geq \delta \\ 0 \qquad for \quad m_i < \delta \\ s.t. \qquad N = \log(\lambda(1 - \delta) + 1) \end{cases} \tag{4}$$

In order to calculated gradient information of images having different exposure values, gamma function is used as proposed by the article. The exposure value is selected based on the following criteria;

$$\text{argmax}_{\gamma}\, M\left(I_{in}^{\gamma}\right) \tag{5}$$

Where M is the sum of gradient information and γ mimic exposure time value. The aim is of this method is to maximize gradient information.

Algorithm 2 is based on third moment of image intensity histogram [36]. Third moment is calculated as follows proposed by the authors as follows;

$$\mu = \sum_{k=0}^{255} (z_k - C)^3\, p(z_k) \tag{6}$$

$$p(z_k) = \frac{q(z_k)}{M*N} \tag{7}$$

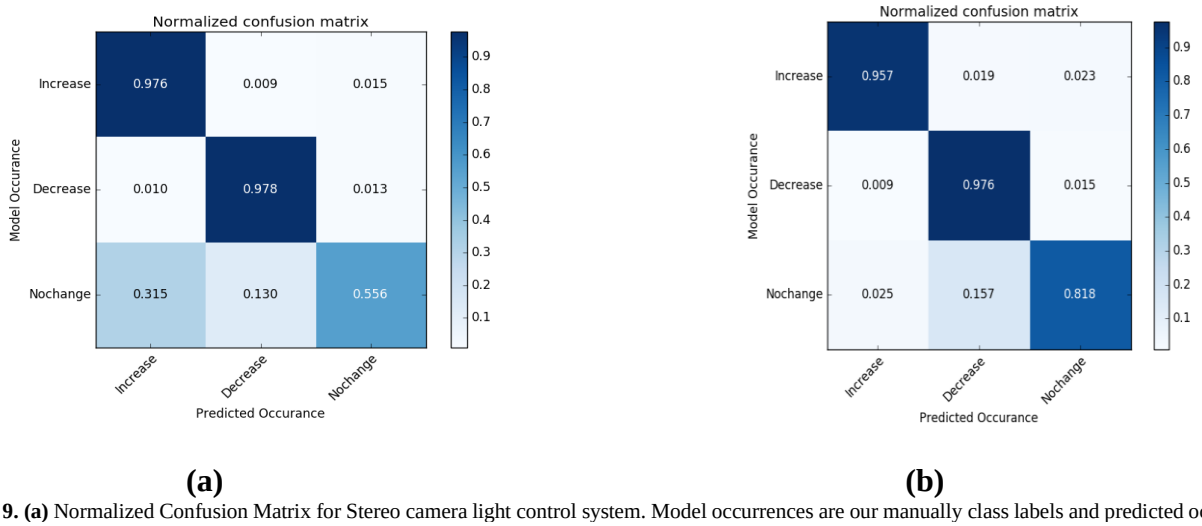
where M*N represent the image size and $q(z_k)$ is the total number of frequencies.

Algorithm 3 is based on middle tone distribution of image intensity histogram [37]. The histogram is divided in to five part and mean sample value (MSV) is used that described as follows;

$$\mu = \sum_{i=0}^{4}(i+1) * x_i / \sum_{i=0}^{4} x_i \quad (8)$$

Where, x_i is the sum of the values in region i and i denotes the region of the histogram. According their paper, the image is correctly exposed when $\mu \approx 2.5$ [37].

Confusion matrices for lighting control in stereo frames and monocular frames is shown in Figure 9. Model occurrences are our manually class label and predicted occurrences are endoscopic illumination controller outcomes. For class intensity increase, according to the normalized confusion matrix it shows 97% accuracy. Most significant error in that case is 0.009% that is decrease light intensity. Similarly, decrease intensity event achieved 97.8% classification accuracy and no change in light intensity event achieved 55.6% classification accuracy. No change light intensity is hard in stereo situation. In our test stereo pair image encountered different illumination in most of situation caused by shadow, surface curvature of internal anatomy etc. Lighting control outcomes for over-exposed, under-exposed and normal exposed classes are presented in Figure 11. Upward arrow, downward arrow and dotted line annotations to stereo images indicate controller output for each case.



(a) **(b)**

Fig. 9. (a) Normalized Confusion Matrix for Stereo camera light control system. Model occurrences are our manually class labels and predicted occurrences are endoscopic illumination controller outcomes for each scene. All the diagonals are true positive. **(b)** Mono camera light control Confusion Matrix. All the diagonals are true positive. Increase light intensity classified by our system with approx. 95% accuracy during the test. Similarly, in approx. 97% situations decrease light intensity events are classified with true positive. Intensity no change events accuracy is about 82%.

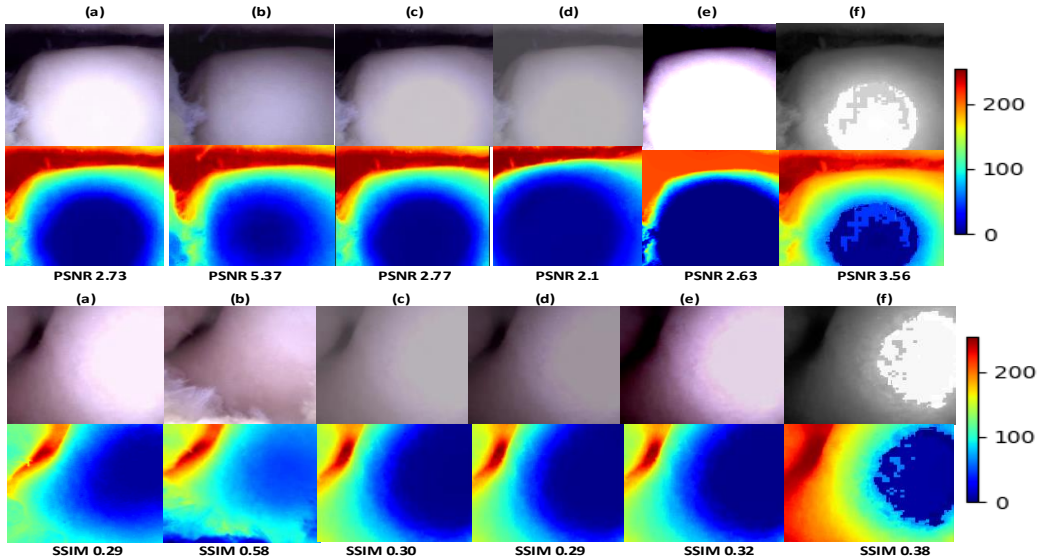


Fig. 10. It compares the performance of our method over pure image post processing-based approach. Column (a) represents the accuracy of the depth map when image is degraded by overexposed pixels. The achieved PSNR and SSIM are 2.73 and 0.29 respectively where well exposed image achieved (b) achieved 5.37 and 0.58. Overexposed image then corrected using correction of brightness (c), decreased contrast (d), increased contrast (e) and through the correction of histogram equalizer(f). The achieved PSNR and SSIM are shows image corrected through the control of illumination provides better solution.

Fig. 11: Light control outcomes for binocular stereo camera on knee arthroscopy video sequences. Each image block contains one stereo image pair, basically left image from left camera and right image from right camera. Image annotations with upward arrow indicates scene requires more light intensity, similarly downward arrows indicate scene contain overexposed portion and need low light intensity and double dashed sign indicates scene meets sufficient illumination condition.

Figure 12. compares the outcome of methods evaluated. Our proposed method received average error of 7.37% and 4.56 % for stereo endoscope and monocular endoscope respectively. All the methods fail to identify no change event for monocular endoscope and except Algorithm 1. It receives 32.7 % error to identify no change error. Where the proposed method is able to detect no change event with 71.27% average accuracy (16.36% in error in stereo and 41.1% error in monocular). Considering overall accuracy to control illumination event the proposed method outperforms state-of-the-art exposure control methods.

EVALUATION OF STEREO IMAGE PAIR				
METHOD	INCREASE LIGHT ERROR (%)	DECREASE LIGHT ERROR (%)	NO CHANGE LIGHT ERROR (%)	AVG. ERROR (%)
ALGO.1	32.7	94.2	32.7	53.2
ALGO.2	0.341	14.25	*N/A	38.19
ALGO.3	25.2	66.34	*N/A	63.84
PROPOSED METHOD	3.4	2.37	16.36	7.37
EVALUATION OF MONOCULAR IMAGE				
METHOD	INCREASE LIGHT ERROR (%)	DECREASE LIGHT ERROR (%)	NO CHANGE LIGHT ERROR (%)	AVG. ERROR (%)
ALGO.1	35.15	88.6	*N/A	74.58
ALGO.2	8	6.4	*N/A	38.13
ALGO.3	22.5	38.5	*N/A	53.66
PROPOSED METHOD	3	1.5	41.1%	4.56

*N/A= Not applicable is considered when error in detecting an event is about 100% approx.

Fig.12: Evaluation of the illumination control methods.

VIII. Conclusion

We have explored the feasibility of support vector machine classifier against the endoscopic camera light intensity control problem in real time. Our results offer good prospects in inferring challenges imposed in surgical scene visualization for MIS into a scene classification problem. Confusion matrices represents the overall performance of our implemented illumination controller. However, accuracy decreases for the 'no change' class category. In knee arthroscopy this event occurs rarely hence very few training data was available for train this category. Additionally, in stereo vision system, two viewpoints receive different level of illumination distribution mostly due to the anatomical structure hence, this specific class category encounters more error. A set of training video sequences under slowly varying light intensity at every possible intensity value is recommended for this approach for further improvement.

Image patch-based similarity index under radiometric, rotation and illumination changes are expected to further enhance feedback in real time. It might be useful to combine patch similarity or dissimilarity index along with the degree of changes provided by content indexing methods. Maximum similarity regions with positive improvement of those reason can infer an intensity event as a true decision, as investigated in this work.

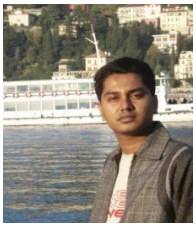
ACKNOWLEDGMENT

This work is supported by Australian Indian Strategic Research Fund Project AISRF53820 and in part by Australian Centre for Robotic Vision.

REFERENCES

1. Jingwei S., Jun W., Liang Z., Shoudong H., Gamini D., "MIS-SLAM: Real-time Large Scale Dense Deformable SLAM System in Minimal Invasive Surgery Based on Heterogeneous Computing", arXiv,March, 2018, arXiv:1803.02009
2. Du, X., Allan, M., Dore, A. et al. “Combined 2D and 3D tracking of surgical instruments for minimally invasive and robotic-assisted surgery”, Int J CARS (2016)
3. R. M. Sousa, M. Wäny, P. Santos and F. Morgado-Dias, “NanEye-An Endoscopy Sensor With 3-D Image Synchronization”, IEEE Sensors Journal, vol. 17,623-631, (2017)
4.Hu M., Penney G., Edwards P., Figl M., Hawkes D.J. (2007) “3D Reconstruction of Internal Organ Surfaces for Minimal Invasive Surgery”. In: Ayache N., Ourselin S., Maeder A. (eds) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2007. MICCAI 2007. Lecture Notes in Computer Science, vol 4791. Springer, Berlin, Heidelberg
5.Collins T., Bartoli A. (2012) “3D Reconstruction in Laparoscopy with Close-Range Photometric Stereo”. In: Ayache N., Delingette H., Golland P., Mori K. (eds) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2012. MICCAI 2012. Lecture Notes in Computer Science, vol 7511. Springer, Berlin, Heidelberg
6. Sørensen, S.M.D., Savran, M.M., Konge, L. et al. “Three-dimensional versus two-dimensional vision in laparoscopy”: a systematic review, Surg Endosc, (2016)
7. Velayutham, V., Fuks, D., Nomi, T. et al. “3D visualization reduces operating time when compared to high-definition 2D in laparoscopic liver resection”: a case-matched study, Surg Endosc, (2016)
8. V. N. Vapnik, "An Overview of Statistical Learning Theory", IEEE Transactions on Neural Networks, vol. 10, no. 5, pp. 988-999, 1999
9. T. Kuno, H. Sugiura and N. Matoba, "A new automatic exposure system for digital still cameras," in IEEE Transactions on Consumer Electronics, vol. 44, no. 1, pp. 192-199, Feb. 1998. doi: 10.1109/30.663747
10. R. Pourreza-Shahri and N. Kehtarnavaz, "Exposure bracketing via automatic exposure selection," 2015 IEEE International Conference on Image Processing (ICIP), Quebec City, QC, 2015, pp. 320-323. doi: 10.1109/ICIP.2015.7350812
11. Seon Joo Kim, J. Frahm and M. Pollefeys, "Radiometric calibration with illumination change for outdoor scene analysis," 2008 IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, 2008, pp. 1-8. doi: 10.1109/CVPR.2008.4587648
12. C. Loscos and K. Jacobs, “High-dynamic range imaging for dynamic scenes,” Computational Photography: Methods and Applications, pp. 259–281, 2010.
13. R. Pourreza-Shahri and N. Kehtarnavaz, "Exposure bracketing via automatic exposure selection," 2015 IEEE International Conference on Image Processing (ICIP), Quebec City, QC, 2015, pp. 320-323.
14.H. Matsushima, “Camera having automatic exposure bracketing device,” Jul. 26 1994, uS Patent 5,333,027.
15.C. Chen, S. Mccloskey and J. Yu, "Analyzing Modern Camera Response Functions," 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), Waikoloa Village, HI, USA, 2019, pp. 1961-1969.
16. R. Janzen and S. Mann, "Feedback Control System for Exposure Optimization in High-Dynamic-Range Multimedia Sensing," 2016 IEEE International Symposium on Multimedia (ISM), San Jose, CA, 2016, pp. 119-125.
17. Raimondo, Schettini, and Corchs Silvia. "Underwater image processing: state of the art of restoration and image enhancement methods." *EURASIP Journal on Advances in Signal Processing* 2010.1 (2010): 746052.
18. Dionysiou A., Agathocleous M., Christodoulou C., Promponas V. (2018) “Convolutional Neural Networks in Combination with Support Vector Machines for Complex Sequential Data Classification”. In: Kůrková V., Manolopoulos Y., Hammer B., Iliadis L., Maglogiannis I. (eds) Artificial Neural Networks and Machine Learning – ICANN 2018. ICANN 2018.
19. Y. H.Geoffroy, K. Sunkavalli, S. Hadap, E. Gambaretto, J.F. Lalonde,"Deep Outdoor Illumination Estimation",arXiv:1611.06403, 2018
20. V. C. Cardei, B. Funt, and K. Barnard, "Estimating the scene illuminant chromaticity by using a neural network," J. Opt. Soc. Am. A 19, 2374-2386 (2002)
21. K. Kawaguchi, L. P. Kaelbling,"Elimination of All Bad Local Minima in Deep Learning", arXiv:1901.00279, 2019
22. I. Shim, T. Oh, J. Lee, J. Choi, D. Choi, I. S. Kweon "Gradient-based Camera Exposure Control for Outdoor Mobile Platforms", arXiv:1708.07338, 2018
23. S. Liu and Y. Zhang, "Detail-preserving underexposed image enhancement via optimal weighted multi-exposure fusion," in IEEE Transactions on Consumer Electronics. 2019,doi: 10.1109/TCE.2019.2893644
24.G. E. Sakr, M. Mokbel, A. Darwich, M. N. Khneisser and A. Hadi, "Comparing deep learning and support vector machines for autonomous waste sorting," 2016 IEEE International Multidisciplinary Conference on Engineering Technology (IMCET), Beirut, 2016, pp. 207-212.
25. Jadhav A.R., Ghontale A.G., Shrivastava V.K. (2019) “Segmentation and Border Detection of Melanoma Lesions Using Convolutional Neural Network and SVM”. In: Verma N., Ghosh A. (eds) Computational Intelligence: Theories, Applications and Future Directions - Volume I. Advances in Intelligent Systems and Computing, vol 798. Springer, Singapore.
26.G. E. Sakr, I. H. Elhajj, H.A.-S. Huijer, "Support vector machines to define and detect agitation transition", Affective Computing IEEE Transactions on, vol. 1, no. 2, pp. 98-108, 2010
27.C. Venkatesan, P. Karthigaikumar, A. Paul, S. Satheeskumaran and R. Kumar, "ECG Signal Preprocessing and SVM Classifier-Based Abnormality Detection in Remote Healthcare Applications," in IEEE Access, vol. 6, pp. 9767-9773, 2018.
28.Qiu, Liang and Ren, Hongliang,"Endoscope Navigation and 3D Reconstruction of Oral Cavity by Visual SLAM With Mitigated Data Scarcity",CVPR, June, 2018
29.Khanam Z., Raheja J.L. (2018) “Tracking of Miniature-Sized Objects in 3D Endoscopic Vision”. In: Das S., Chaki N. (eds) Algorithms and Applications. Smart Innovation, Systems and Technologies, vol 88. Springer, Singapore
30. Pedregosa, F. et. al,"Scikit-learn: Machine Learning in Python", Journal of Machine Learning Research,vol.12, pp 2825-2830, 2011
31.Zabih, R., Woodfill, J.:”Non-parametric local transforms for computing visual correspondence”. In: Eklundh, J.-O. (ed.) ECCV 1994. LNCS, vol. 801, pp. 151–158. Springer, Heidelberg (1994)
32.Hafner D., Demetz O., Weickert J. (2013) “Why Is the Census Transform Good for Robust Optic Flow Computation?”. In: Kuijper A., Bredies K., Pock T., Bischof H. (eds) Scale Space and Variational Methods in Computer Vision. SSVM 2013. Lecture Notes in Computer Science, vol 7893. Springer, Berlin, Heidelberg
33.Zhou Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," in IEEE Transactions on Image Processing, vol. 13, no. 4, pp. 600-612, April 2004.
doi: 10.1109/TIP.2003.819861
34. J. Lim, Y. Kim and S. Lee, "A census transform-based robust stereo matching under radiometric changes," *2016* Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA), Jeju, 2016, pp. 1-4.
35 I. Shim, J. Lee and I. S. Kweon, "Auto-adjusting camera exposure for outdoor robotics using gradient information," 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems, Chicago, IL, 2014, pp. 1011-1017
36. Nourani-Vatani, Navid, and Jonathan M. Roberts. "Automatic camera exposure control." *Proceedings of the Australasian Conference on Robotics and Automation 2007*. Australian Robotics & Automation Association ARAA, 2007..
37. Wen Chen and Xinglong Li (2017) “Exposure Evaluation Method Based on Histogram Statistics” 2nd International Conference on Electrical, Automation and Mechanical Engineering (EAME 2017)
38. D. Xu, C. Doutre and P. Nasiopoulos, "Saturated-pixel enhancement for color images," 2010 IEEE International Symposium on Circuits and Systems (ISCAS), Paris, 2010, pp. 3377-3380. doi: 10.1109/ISCAS.2010.5537871
39. D. Xu, C. Doutre and P. Nasiopoulos, "Correction of Clipped Pixels in Color Images," in IEEE Transactions on Visualization and Computer Graphics, vol. 17, no. 3, pp. 333-344, March 2011. doi: 10.1109/TVCG.2010.63

40. Masood, S. Z., Zhu, J. and Tappen, M. F. (2009), "Automatic Correction of Saturated Regions in Photographs using Cross-Channel Correlation. Computer Graphics Forum", 28: 1861-1869. doi:10.1111/j.1467-8659.2009.01564.x.
41. Masood, S. Z., Zhu, J. and Tappen, M. F. (2009), "Automatic Correction of Saturated Regions in Photographs using Cross-Channel Correlation. Computer Graphics Forum", 28: 1861-1869. doi:10.1111/j.1467-8659.2009.01564.x.
42. Rodríguez-Quiñonez, J. C., et al. "Improve a 3D distance measurement accuracy in stereo vision systems using optimization methods' approach." *Opto- Electronics Review* 25.1 (2017): 24-32.
43. Ramírez-Hernández, Luis R., et al. "Improve three-dimensional point localization accuracy in stereo vision systems using a novel camera calibration method." *International Journal of Advanced Robotic Systems* 17.1 (2020): 1729881419896717.
44. Rodríguez-Quiñonez, J. C., et al. "Improve 3D laser scanner measurements accuracy using a FFBP neural network with Widrow-Hoff weight/bias learning function." *Opto-Electronics Review* 22.4 (2014): 224-235.
45. Rodriguez-Quinonez, Julio C., et al. "Surface recognition improvement in 3D medical laser scanner using Levenberg–Marquardt method." *Signal Processing* 93.2 (2013): 378-386.
46. Miranda-Vega, Jesús E., Moisés Rivas-López, and Wendy Flores-Fuentes. "k-Nearest Neighbor Classification for Pattern Recognition of a Reference Source Light for Machine Vision System." *IEEE Sensors Journal* (2020).
47. Miranda-Vega, Jesus E., et al. "Digital Implementation of FIR Filters for the Minimize of Optical and Optoelectronic Noise of an Optical Scanning System." *RIAI-REVISTA IBEROAMERICANA DE AUTOMATICA E INFORMATICA INDUSTRIAL* 16.3 (2019): 344-357.
48. Shahnewaz, Ali, and Ajay K. Pandey. "Color and Depth Sensing Sensor Technologies for Robotics and Machine Vision." *Machine Vision and Navigation*. Springer, Cham, 2020. 59-86.
49. Jansen-van Vuuren, Ross D., Ali Shahnewaz, and Ajay K. Pandey. "Image and Signal Sensors for Computing and Machine Vision: Developments to Meet Future Needs." *Machine Vision and Navigation*. Springer, Cham, 2020. 3-32.
50. Jonmohamadi, Yaqub, et al. "Automatic segmentation of multiple structures in knee arthroscopy using deep learning." IEEE Access 8 (2020): 51853-51861.
51. Flores-Fuentes, Wendy, et al. "Combined application of power spectrum centroid and support vector machines for measurement improvement in optical scanning systems." Signal Processing 98 (2014): 37-51.
52. Zainuddin, AZ Ahmad, et al. "Comparison Between Support Vector Machine with Polynomial and RBF Kernels Performance in Recognizing EEG Signals of Dyslexic Children." World Congress on Medical Physics and Biomedical Engineering 2018. Springer, Singapore, 2019.
53. Savas, Caner, and Fabio Dovis. "The Impact of Different Kernel Functions on the Performance of Scintillation Detection Based on Support Vector Machines." *Sensors* 19.23 (2019): 5219.



Mr Shahnewaz Ali is a Ph.D. researcher at Queensland University of Technology, Australia and undertaking his research on medical robotics. His current research is addressing some of the challenges to robotic vision whithin the body more especifically within knee that includes perception and awareness as well as improve technological support to the imaging system.

Dr Yaqub Jonmohamadi obtained his Ph.D. title in neuroimaging, with focus on EEG analysis, from the University of Otago, New Zealand, in 2014. He started his first post-doctoral fellowship in multi-modal neuroimaging at the University of Auckland, New Zealand in 2015. Currently, he is undertaking his second post-doctoral fellowship in medical robotics and medical imaging at the Queensland University of Technology, Australia.

Dr Yu Takeda is an orthopaedic surgeon. He studied medicine at the Hyogo College of Medicine (Japan) between 2003 and 2009 and was awarded a Ph.D. degree by the Hyogo College of Medicine in 2018. He is currently a visiting researcher at the Queensland University of Technology (Australia) in the field of robot assisted and autonomous surgery.

Prof Jonathan Roberts is a Professor in Robotics at Queensland University of Technology (QUT), is a Chief Investigator at the Australian Centre for Robotic Vision (ACRV) and is currently QUT's Robotics and Autonomous Systems Discipline Leader. His main research interests are in the areas of Field Robotics, Medical Robotics, Performance Robotics and more recently Design Robotics.

Prof Ross Crawford holds the position of Chair in Orthopedic Research at Queensland University of Technology. He is an internationally recognized expert in orthopedics and robotic assisted surgery.

Dr Ajay Pandey is a Senior Lecturer in Robotics and Autonomous Systems at the School of Electrical Engineering and Robotics. His research interest encompasses application of Photonics and Molecular Electronics for designing intelligent vision and tactile sensors for medicine and robotics.