International Journal of Medical, Medicine and Health Sciences
ISSN: 2517-9969
Vol:7, No:3, 2013

The Visual Inspection of Surgical Tasks Using Machine Vision: Applications to Robotic Surgery

M. Ovinis, D. Kerr, K. Bouazza-Marouf, M. Vloeberghs

***Abstract*—**In this paper, the feasibility of using machine vision to assess task completion in a surgical intervention is investigated, with the aim of incorporating vision based inspection in robotic surgery systems. The visually rich operative field presents a good environment for the development of automated visual inspection techniques in these systems, for a more comprehensive approach when performing a surgical task. As a proof of concept, machine vision techniques were used to distinguish the two possible outcomes i.e. satisfactory or unsatisfactory, of three primary surgical tasks involved in creating a burr hole in the skull, namely incision, retraction, and drilling. Encouraging results were obtained for the three tasks under consideration, which has been demonstrated by experiments on cadaveric pig heads. These findings are suggestive for the potential use of machine vision to validate successful task completion in robotic surgery systems. Finally, the potential of using machine vision in the operating theatre, and the challenges that must be addressed, are identified and discussed.

***Keywords*—**Machine vision, robotic surgery, visual inspection.

I. Introduction

A surgical task is a single action that is performed during a surgical procedure, which is usually made up of several surgical actions performed sequentially. The ability to assess the outcome of a surgical action is an important one, as subsequent surgical actions cannot usually be performed prior to the successful completion of the previous surgical action. This capability is therefore essential from both a safety and operational point of view in robotic surgery systems. An important distinction to be made in the use of different sensory capabilities in these systems is in terms of their purpose. While force, torque and positional sensors are typically used to perform surgical tasks, a visual sensor would be used to assess its completion. In principle, sensors used to validate the completion of a task should be independent from those used to perform the task being assessed, to avoid a situation where errors, if any, become compounded. Incorporating visual sensory capability therefore improves the reliability and robustness of robotic surgery systems, which traditionally only employ force, torque and positional measurements [1]. A visual sensor can often provide information where other sensors cannot. These different sensory capabilities complement each other, allowing for a more comprehensive approach to the automation of a surgical task.

M. Ovinis is a lecturer in the Department of Mechanical Engineering, Universiti Teknologi PETRONAS, 31750, Tronoh, Malaysia (phone: +6053687150; fax: +6053656461; email: mark_ovinis@petronas.com.my).

D. Kerr and K. Bouazza-Marouf are Senior Lecturers in the Department of Mechanical and Manufacturing Engineering, Loughborough University, LE11 3TU, Loughborough, UK (e-mail: d.kerr@lboro.ac.uk and k.bouazza-marouf@lboro.ac.uk, respectively).

Prof. M. Vloeberghs is a Consultant Neurosurgeon at the Queen's Medical Centre, Nottingham University, NG7 2UH, Nottingham, UK (e-mail: Michael.Vloeberghs@Nottingham.ac.uk).

II. Previous Work

Closely related to this work is the use of computer vision for the diagnosis of malignant melanoma, where the subject of the inspection is biological specimens. Encouragingly, diagnostic accuracy rivaling those achieved by experts has been reported [2], [3], an impressive feat considering 1) the clinical signs of malignancy are often ambiguous [4] and 2) the use of low-level features only such as asymmetry, color, texture and size in the inspection. Other uses of vision based inspection related to biological specimens include the automatic monitoring and analysis of human embryonic stem cell growth, by classifying their morphologic changes based on features extracted from their optical images [5]. Automating the inspection proved a reliable means of obtaining more data compared to conventional methods, as well as allowing the continuous monitoring of these cells without disturbing them.

While machine vision is routinely employed in industrial settings for automated inspection tasks, its use in other less conventional settings such as in the operating theatre has been less widespread. Lo et al. [6] developed a framework for the classification of surgical episodes using multiple visual cues related to shape, deformation, changes in light reflection and other low level image features. Specifically, their work was on the detection of surgical actions, such as the interaction of surgical instruments with soft tissue, retraction, cauterization, and suturing, in video sequences of minimally invasive laparoscopic surgeries. Padoy et al. [7] developed an approach to recognize a subset of surgical actions performed by the surgeon during laparoscopic surgery, using visual cues and signals recorded from surgical instruments, to monitor the progress of an operation. The objective was to use the information gathered to design a support system for an operating theatre e.g. activation of the operating theatre lights, automatic reporting, etc. The examples on the use of computer vision in the operating theatre are primarily concerned with the detection of surgical tasks. A natural progression would be the inspection of the outcome of these tasks. To the authors' knowledge, the use of machine vision in this context has not been reported.

III. Methodology

To illustrate the concept, practical examples of the visual inspection of the outcome of surgical tasks to create a burr

hole in the skull were used. While a burr hole is a relatively simple procedure (when performed by a neurosurgeon), it is very complex to perform robotically. A subset of this procedure is to determine the successful completion of a given surgical task, to validate its completion. Towards this end, the use visual cues were investigated. Burr holes were created on cadaveric pig heads to simulate the two possible outcomes i.e. complete/satisfactory and incomplete/unsatisfactory, of the three primary surgical tasks involved in making the hole, and to acquire the corresponding images. Practical and ethical considerations preclude experimentation on humans or live animals. As such, animal cadavers were used as a substitute. In terms of the selection of an animal cadaver, pig heads were used because of their anatomical similarity to humans. With the exception of primates, the anatomy of the temporoparietal region (site of a burr hole) of pigs most closely resembles that of a human [8]. As access to the pig's brain is easier through the frontoparietal region (due to its proximity to the brain and structure of the pig's skull), this region was chosen as the site of the burr hole, instead of the temporoparietal region.

The surgical procedure for creating a burr hole in a pig, as described by Kaiser and Fruhauf [9], is similar in many respects to that for creating a burr hole in a human and was adopted with some modifications. A midline incision, approximately 3cm long, centered over the site of the burr hole, was made with a #10 scalpel blade, at different depths up to the pericranium to simulate incomplete incisions, and a single incision through the layers of the scalp to simulate complete incisions. Incisions are said to be incomplete if all the layers of the scalp including the pericranium have not been incised, and are classified as complete if all the layers of the scalp including the pericranium have been incised. The complete incisions were subsequently retracted at varying degrees using a self-retaining retractor to expose the cranium, until the incision was wide enough to accommodate a burr hole. A burr hole 10mm in diameter was then drilled in the cranium using a cordless drill at high speed. Over fifty experiments were performed to obtain a representative sample, owing to the variability associated with biological subjects. Images of these outcomes were analyzed to identify appropriate visual cues and characteristic features to distinguish the two possible outcomes of each surgical task.

A. Experimental Setup

A charge-coupled device (CCD) video camera, with a white light emitting diode (LED) ring light mounted around its optical lens, was used to obtain video images of the surgical tasks. The ring light used had an adjustable intensity and a four-section quadrant control, enabling each quadrant to be individually controlled. Both the camera focus and exposure settings were set to automatic. To facilitate the placement of retractors, the incision is usually pulled apart by applying tension parallel to the plane of the skin. To simulate this parting of the skin, two square tabs (29x29mm) were attached to the skin with cyanoacrylate adhesive [10] and pulled apart by two 600gram weights using a pulley and weight system (see Fig. 1). The distance between the tabs was 29mm.

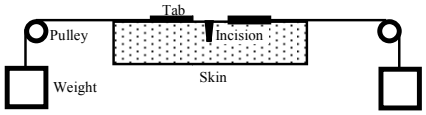


Fig. 1 Schematic of the parting of the skin

B. Characterization of an Incision

The first illustrative example is the inspection of skin incisions. When surgeons incise the scalp, the scalpel blade is pushed into the skin, through the layers of the scalp, until it has touched the cranium. A complete incision should be both sufficiently long and deep i.e. an incision up to the pericranium along the length of the incision. In this work, machine vision was used to inspect an incision for sufficient depth i.e. an incision that was incised up to the cranium. The incisions were created such that they were of the required length, the only variable being the depth of incision. It is assumed that the position, length and orientation of the incision are satisfactory, although these properties could be determined as well using machine vision techniques. For example, assuming that the incision can be segmented from the image and that the scale of the image is known, the properties of an image region corresponding to the incision such as position, length and orientation can be measured. The position of the incision is given by the centroid or the centre of mass of the image region. The length of the incision is the Euclidean distance between the two left and right extreme points of the image region. Its orientation (relative to the camera axis) is the angle between the horizontal axis and the major axis of the ellipse that has the same second-moments as the image region. Two types of incisions, complete and incomplete, were created (see Fig. 2).

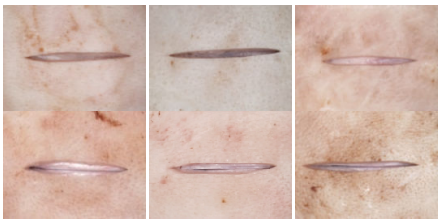


Fig. 2 Incomplete incision (top) and complete incision (bottom)

Complete incisions are incisions incised through all the layers of the skin up to the cranium. Incisions up to the pericranium were considered incomplete. These images were subsequently analyzed to determine if a distinction could be made between complete and incomplete incisions based on their image properties. It was observed that complete incisions exhibit a characteristic dark line, made as the scalpel blade incises the skin up to the cranium, which is not present in incomplete incisions. As such, the presence of a dark line along the length of an incision might be used to identify a complete incision. To detect the presence of the dark line corresponding to a complete incision, the line intensity profiles of six equally spaced, 13-pixel wide line segments across the incision were determined (see Fig. 3).

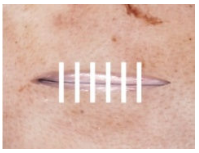


Fig. 3 Line segments across an incision (drawn to scale)

This method does not require that the incision be segmented from the image. However, the start and end position of the incision is required. As the incision would be made robotically, it is assumed that these positions would be known. A pixel on the incision line (characteristic dark line indicative of a complete incision) corresponds to a pixel along the line segment where the second derivative of the line intensity profile is maximum and exceeds an empirically determined threshold of 0.015 (see Fig. 4).

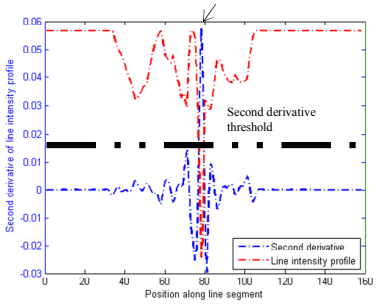


Fig. 4 Intensity profile of a line segment across an incision and its second derivative. The position of the pixel on the incision line along the line segment is denoted by an arrow

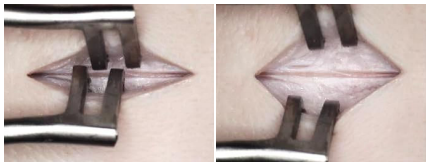
Because techniques based on line intensity profiles and derivatives are susceptible to noise, several noise reduction techniques were used. First, the image was filtered with a low pass 3-by-3 Gaussian filter with a standard deviation of 0.5. The mean value of the 13-pixel wide line intensity profile of each line segment was used, and fitted to a cubic spline, to reduce the effects of any outliers by way of smoothing. The use of a relatively wide line segment also reduces the potential of erroneously detecting the edges of the incision, as the line intensity profile of the incision line would be closer to the profile of a complete incision line than that of the profile at the curved edges of an incision. To reduce the effects of specular reflection, which distorts second derivative values of the line intensity profile, the greyscale intensity values of pixels along a line segment greater than the mean of the line intensity profile was replaced with its mean value, based on the following equation:

$$I(x) = \begin{cases} \bar{I} & I(x) > \bar{I} \\ I(x) & I(x) \leq \bar{I} \end{cases} \quad (1)$$

where $I(x)$ is the greyscale intensity value of a pixel along the line segment at position x and $\bar{I}$ is the mean greyscale intensity value of all pixels along the line segment. Additionally, techniques such as temporal filtering could be used where video images are available. However, in this work, only still images of the incision were used.

C. Characterization of a Retracted Incision

An incision is considered completely retracted if it is able to accommodate a burr hole of a given size. While the retraction of an incision may be determined indirectly from the separation distance of the prongs of the retractors (see Fig. 5), a potential problem with this approach is that the retractors may have extended without the incision being retracted. This would be the case if the prongs of the retractor have not engaged the edges of the incision, leading to the erroneous assumption that the incision is sufficiently retracted.



(a) (b)

Fig. 5 (a) Incomplete retraction and (b) complete retraction

To assess if an incision has been sufficiently retracted, the maximum circular free space within the retracted incision was determined by fitting the largest possible circle inside a region bounded by the edges of the incision and the prongs of the retractor (see Fig. 6). If the diameter of this circle is larger than the diameter of the prospective burr hole (within a specified tolerance), the incision is considered sufficiently retracted. The actual diameter of the fitted circle can be found if the scale of the image is known. A possible solution to determine image scale is by using the pixel separation distance of the prongs on each side of the retractor as a guide. Assuming that the actual distance is known a-priori, the scale of the image and hence the actual diameter of the fitted circle can be found.

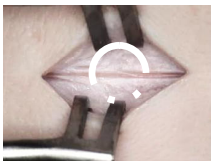


Fig. 6 The largest circular free space area within the retracted incision

To define a region bounded by the edges of the incision and retractors, these edges would first have to be detected in the image. As the retractors are made from surgical stainless steel and have a dull grey color that is distinct in the operative site, color-based segmentation using k-means clustering was used. The number of clusters to be partitioned was specified automatically based on the number of local maxima of a three-dimensional color histogram, which corresponds to the number of regions with the most dominant color in the image. The edges of the retracted incision were segmented using a Canny edge detector. However, these edges cannot always be adequately detected because of the sometimes low contrast at

the boundaries of the incision. To improve edge detection, sections of the ring light were turned off to cast shadows at the boundaries of the incision to increase their contrast. Fig. 7 illustrates the technique.

Fig. 7 (From left to right) Detected edges of a retracted incision by alternating each quadrant of the ring light. The last figure shows the composite image

Once the edges of the incision and retractors were detected, a distance map based on the distance transform of all the pixels in the image to an edge map of the retractors and the retracted incision was determined. The pixel with the maximum distance within a circular region of interest is the centre of a circle that corresponds to the maximum free circular space. The radius of the fitted circle is the distance of this pixel to either the nearest edge pixel of the prongs of the retractors or the edge pixel of the retracted incision. Fig. 8 illustrates the technique. It is important that pixels within the region bounded by the edges of the incision and retractors be removed, as the position and diameter of the fitted circle will be affected. These pixels were removed by selecting the first and last pixel when traversing the edge map horizontally and vertically.

Fig. 8 (Left to right) An edge map, a distance map overlaid with the edge map and a circular region of interest, and the fitted circle

D. Characterization of a Burr Hole

The inspection of a burr hole is the most straightforward of the three surgical tasks under consideration. Fig. 9 shows an incomplete and complete burr hole. An incomplete or partial burr hole is where the skull has not been penetrated. A complete burr hole is one where the underlying dura mater i.e. the outermost part of the brain, is visible. The dura mater is primarily grayish in color and has a membrane like appearance and texture. Depending on whether the partial burr hole has been drilled to the cortical or cancellous layer of the skull, it would have either a smooth or a 'spongy' appearance, although both will be predominantly white. Possible features to distinguish the two types of burr holes would therefore include color and texture.

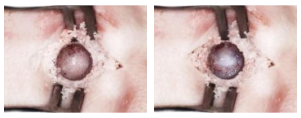


Fig. 9 Incomplete/partial burr hole (left) and complete burr hole (right)

For the visual inspection of a burr hole, it is first necessary to detect its presence in an image. Because of its circular shape, a Hough transform to detect circles was used for the detection of a burr hole. A circle is disregarded if the ratio of its accumulator cell count i.e. the number of detected edge pixels, to the circle perimeter is less than 10%. This ratio eliminates false positives i.e. a burr hole is sometimes 'detected' when there is none. The 10% ratio was found to represent an acceptable compromise between false detection and under detection. The circle with the highest ratio of detected edge pixels i.e. accumulator cell count, to the number of pixels along the perimeter of the circle (calculated based on the radius of the circle), was selected. This is so that the algorithm was not biased towards larger circles, which may have a higher accumulator count over smaller circles. Fig. 10 shows examples of the detected burr holes using this technique.

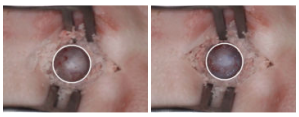


Fig. 10 Detected burr holes

The CIE delta-E color difference, ΔE_{ab*} between a burr hole and a surrounding circular area, with a radius of 1.5 times the radius of the hole (see Fig. 11) was subsequently computed to determine if color difference could be used as a basis to differentiate incomplete and complete burr holes:

$$\Delta E_{ab*} = \sqrt{(L_2^* - L_1^*)^2 + (a_2^* - a_1^*)^2 + (b_2^* - b_1^*)^2} \qquad (2)$$

where L^*, a^* and b^* are the three coordinates of the CIELAB color space and ΔE_{ab*} is the Euclidean distance between two colors in CIELAB space. A $\Delta E_{ab*} \approx 2.3$ corresponds to a just noticeable difference [11], below which the two colors are indistinguishable (unless placed adjacent to one another). Textural information to differentiate incomplete and complete burr holes was also considered, as incomplete and complete burr hole are predominantly smooth and rough (in terms of their grey levels representation) respectively.

Fig. 11 Circular areas used for the determination of ΔE_{ab*}

IV. Results and Discussion

A. Incision

Fig. 12 shows illustrative examples of the output of the incision detection method, based on the detection of its line intensity profile, for complete and incomplete incisions. The crosses represent the pixels that correspond to the perceived incision (complete).

Fig. 12 Illustrative examples of the output of the complete incision detection method for complete and incomplete incisions

Sixty-four images consisting of 27 incomplete and 37 complete incisions were analyzed. Fig. 13 shows the number of lines segments where an incision line is detected, for complete and incomplete incisions. Incomplete and complete incisions can be partially differentiated based on the number of line segments with an incision line detected. An incision is classified as incomplete if three or less line segments where an incision line was detected were found.

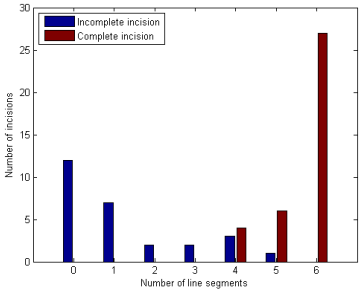


Fig. 13 The number of line segments with a complete incision line detected, for incomplete and complete incisions

The mean residual of a best-fit line (in the least squares sense) fitted to the pixels corresponding to the perceived incision line (see Fig. 14) was determined and used to differentiate incomplete and complete incisions further. To account for the difference in scales in the images, the ratio of the residual to the distance between the line segments was used.

Fig. 14 Line fitted to the pixels corresponding to the perceived incision line

The mean residual of this fitted line for incomplete incisions (with an incision line detected in four or more line segments) and complete incisions are shown in Figs. 15 and 16.

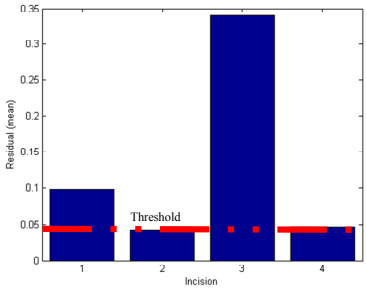


Fig. 15 Mean residuals for incomplete incisions (with an incision line detected in four or more line segments). The red dotted line is the mean residual threshold above which an incision is considered incomplete

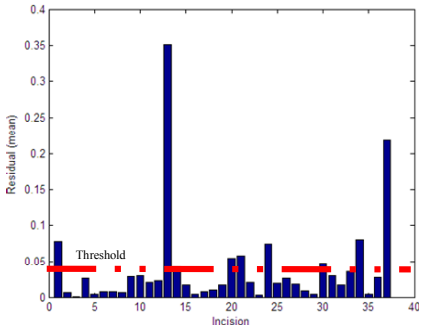


Fig. 16 Mean residuals for complete incisions. The red dotted line is the mean residual threshold in Fig. 15

By using a mean residual threshold of 0.04 for an incision to be considered as a complete incision, all incomplete incisions (in this work) can be precluded, although the overall misclassification error rate would increase. Nine out of the 64 sample incisions were misclassified, a misclassification error rate of 14.1%. Crucially however, is that all the misclassifications are false negatives i.e. complete incisions classified as incomplete. In the visual inspection of an incision, false negatives i.e. complete incisions classified as incomplete are preferable to false positives i.e. incomplete incisions classified as complete. Erroneously classifying an incomplete incision as complete will result in an attempt by the robotic system to retract the incision. On the other hand, a false negative would only require further inspection before the surgical procedure could be continued, a minor inconvenience compared to the possible complications that may arise by attempting to retract an incision that has not been properly incised. Misclassifications of complete incisions were due to the mean residual of the fitted line of these incisions exceeding the threshold of 0.04 (selected to avoid misclassification of incomplete incisions). The high mean residual of misclassified complete incisions were due to pixels being incorrectly identified as belonging to an incision line. This is usually due to the incision line having a lower contrast compared to the edges/boundaries of the incision (see Fig. 17), as This is one of the problems with this approach of classifying incisions, The maximum second derivative of the

line intensity profile does not always correspond to the incision line (see Fig. 18). A possible solution would be to assign greater weightage to the line intensity profile closer to the middle of the line segment, to improve the detection of an incision line over the edges of an incision. However, this technique may not work if the incision line deviates from the middle of the line segment.

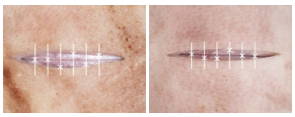


Fig. 17 Examples of misclassified complete incisions

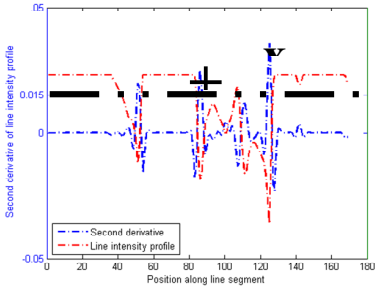


Fig. 18 Intensity profile of a misclassified complete incision. 'x' indicates the perceived position of the incision line and '+' indicates its actual position

B. Retraction

Fig. 19 illustrates examples where the largest circle within the retracted incision was successfully fitted and examples where it failed. In all examples where the technique failed, this was because the edges of the retracted incision or the prongs of the retractor were not segmented well. The segmentation of the retractors failed in some images because the number of clusters (determined automatically) used for color-based segmentation was inadequate. As a result, due to the proximity of the color of the retractor with part of the background skin and/or shadows, the retractors are sometimes segmented along with these regions. Segmenting the retractors robustly is difficult because of the different skin/background colors, the presence of shadows and specular reflections. Specular reflections, due to the reflectivity of the retractor surface, can be reduced by using retractors with a matt surface (instead of retractors with a mirror like surface) as well as by using a diffuser to provide softer illumination of the scene. This also reduces shadows in the scene. A possible solution is to contrive the color segmentation and recognition of the retractors to be simple by using a retractor with a distinct color that will not be mistaken with its background and is not present in the operative field, such as the color cyan. However, this would involve either the use of non-standard retractors, or painting the retractors (which may not be feasible in terms of sterility requirements). Although the edges of the incision were not always detected in the majority of the images (a directional lightning was not used to facilitate detection of the edges), this technique has successfully fitted relatively large circles within the retracted incision. The technique is therefore relatively robust to missing edges and does not always require that all the edges of the incision be found. Nevertheless, as shown in Fig. 19, where edges of the incision are not adequately detected, the position and size of the fitted circle can be erroneous. Where the edges of the retracted incision were not properly detected, the use of directional lightning would facilitate its detection (see Fig. 7).

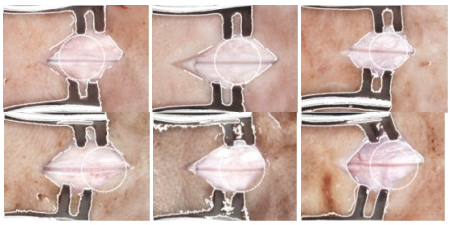


Fig. 19 Illustrative examples of a correctly (top) and incorrectly (bottom) fitted circle within the retracted incision

C. Burr Hole

A leave-one-out cross validation using three classifiers, naïve Bayes, LDA and QDA, was performed to classify a random subset of 54 images of partial burr holes and 54 images of complete burr holes from a total of 162 images (54 images of complete burr holes and 108 images of partial burr holes that were labeled for ground-truth). These sRGB images of burr holes were first converted into the CIELAB color space (based on ITU-R Recommendation BT.709 using the CIE illuminant D50 white point reference). Each image from the dataset was used as validation data, and the remaining images used as training data. This process is repeated such that each image is used once as the validation data. The mean and maximum classification error were obtained by repeating the training and testing of the images ten times. The leave one out naïve Bayes, LDA and QDA misclassification error rate based on the ΔE_{ab*} of incomplete and complete burr holes are shown in Table I.

TABLE I
MEAN AND MAXIMUM MISCLASSIFICATION ERROR RATE (%) BASED ON THE ΔE_{ab*} FOR INCOMPLETE AND COMPLETE BURR HOLES

Classifier		
Naïve Bayes	LDA	QDA
19.5 / 24.1	19.5 / 24.1	18.7 / 22.2

Each burr hole was subsequently represented by the mean, median and mode pixel CIELAB color components of all its pixels. Table II shows the leave one out LDA misclassification error rates based on the color component values of incomplete and complete burr holes (classification results for naïve Bayes and QDA not shown).

TABLE II
LDA LEAVE ONE OUT MISCLASSIFICATION ERROR RATE (%)

LDA leave one out misclassification error rate (%)							
Statistical Measure	Colour space						
	L*	a*	b*	L*a*	L*b*	a*b*	L*a*b*
Mean	13.8	35.2	10.8	8.9	3.7	10.7	2.3
(average/maximum)	14.8	38.9	13.0	10.2	6.5	13.9	3.7
Median	13.3	36.6	13.2	7.0	4.1	13.1	4.1
(average/maximum)	15.7	39.8	15.7	10.2	5.6	15.7	4.6
Mode	14.4	34.8	19.3	9.8	6.1	19.4	6.0
(average/maximum)	18.5	37.0	21.3	13.0	9.3	21.3	9.3

The best misclassification error rate is 2.3% with a maximum error rate of 3.7% (for 10 iterations) based on mean L*a*b* values using LDA. This good classification performance can be attributed to the good separability of the two types of burr hole in CIELAB color space. Fig. 20 shows the separation of the mean pixel values of all incomplete and complete burr holes in CIELAB color space. The misclassification error rate i.e. the percentage of samples that are misclassified weighted by the prior probabilities for the group, was 1.9%. The mean ΔE_{ab*} between the two classes is 34.4, which is considerably greater than the just noticeable difference $\Delta E_{ab*} \approx 2.3$.

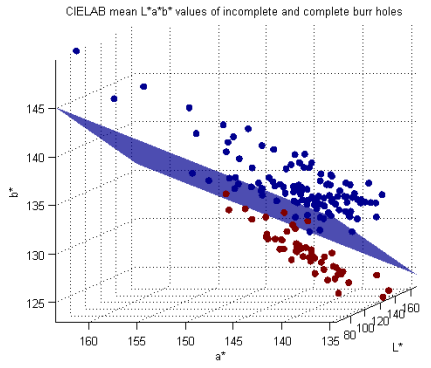


Fig. 20 Scatterplot of pixels values (mean) of incomplete and complete burr holes in CIELAB color space and the LDA decision boundary separating the two classes

In general, all the classifiers performed equally. Although these classifiers provided good classification performance, in this work, a false positive i.e. partial burr holes incorrectly classified as a complete burr hole would result in the subsequent surgical task being performed. Classification performance e.g. sensitivity and specificity can be changed using a different decision criterion threshold value. Geometrically, this threshold value corresponds to the location of the decision boundary i.e. the plane surface in Fig. 20. A Receiver Operating Characteristic curve could be used to identify threshold values such that there are no false positives. Although this will inevitably reduce the sensitivity (number of true positives correctly identified) of the inspection and increase the amount of false negatives (complete burr holes misclassifies as partial burr hole), a higher overall misclassification rate without any false positives is preferable to a lower misclassification error rate with false positives.

In terms of textural information, four of Haralick's coefficients [12] for a grey-level co-occurrence matrix (GLCM), contrast, correlation, energy and homogeneity, were computed and used as features for classification. A k x k x 8 GLCM was used (k = 8,16,32,64,128 and 256), with a distance of one pixel in each direction for a pixel of interest and its corresponding pair/neighbor. Classification was performed using all permutations of the different texture measures. The maximum misclassification error rate ranges from 13.0 – 13.9% with the average misclassification error rate ranging from 9.7 – 12.0%. The lowest leave one out misclassification error rate was obtained using LDA based on all four texture measures i.e. contrast, correlation, energy and homogeneity, with a mean misclassification error rate of 9.7% and a maximum error rate of 13.0%. This misclassification error rates is considerably higher than the mean misclassification error rate of 2.3% and a maximum error rate of 3.7% achieved using color information alone. Although texture is a weaker feature in discriminating the two classes of burr holes compared to color, the combined use of texture and color was considered. Table III shows the LDA leave one out misclassification error rates based on mean color values in CIELAB color space and texture properties of incomplete and complete burr holes

TABLE III
MISCLASSIFICATION ERROR RATE (%) BASED ON COLOR AND TEXTURE PROPERTIES OF INCOMPLETE AND COMPLETE BURR HOLES

LDA leave one out misclassification error rate (%)											
Feature Space	Iteration										Average/ Maximum
Colour	3.7	3.7	1.9	1.9	3.7	2.8	3.7	3.7	1.9	2.8	2.3 3.7
Colour + Texture	3.7	2.8	2.8	3.7	3.7	0.9	2.8	2.8	2.8	3.7	3.0 3.7

The average misclassification error rate based on combined color and texture information is 3.0% with a maximum misclassification error rate of 3.7%. On the other hand, the lowest average and maximum misclassification error rate are 2.3% and 3.7% respectively, using color information only. Combining textural information with color information did not change classification performance in terms of the maximum misclassification error rate. Fig. 21 shows the images that the classifier could not discriminate based on color and texture alone. All misclassified burr holes were false positives i.e. partial burr holes incorrectly classified as a complete burr hole. Further testing with a greater number of sample images is needed to determine if the use of texture measures will aid in classification performance.

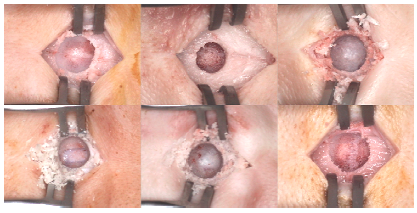


Fig. 21 Misclassified burr hole

Although the images from the CCD video camera are in sRGB color space, a device independent color model, it is not known if the images acquired are color-accurate sRGB because the CCD was not calibrated for color. Nevertheless, although the color values are not absolute colorimetric values, assuming similar color calibration for all images, this work has shown that the two types of burr holes can be separated in terms of their color. In terms of the best feature to represent a burr hole, its mean pixel color value (in CIELAB color space) gave the best misclassification error rate, with a better classification performance compared to using color difference. In terms of classifier performance, the misclassification error rates were comparable for naïve Bayes classifier, LDA and QDA, with LDA performing slightly better, suggesting that the feature space is linearly separable. Although different classifiers such as neural networks might give better classification results, because of the relatively small dataset, the emphasis was on finding appropriate features to perform the classification rather than strength of the classifiers.

V. Conclusion

A vision-based approach to inspect three primary surgical tasks for the creation of a burr-hole was presented. To validate the approach, experiments were performed on cadaveric pig heads, to simulate different fault scenarios, and to assess completion by using visual cues from the scene. The use of machine vision to inspect the outcome of surgical tasks has never been considered before. The classification rate for incisions was 85.9% while classification rate for burr holes was 97.7%. The good burr hole classification rate can be attributed to incomplete and complete burr holes having characteristic colors, and these being consistent through a wide range of subjects. Classification simply becomes a matter of discriminating the two classes based on their characteristic colors.

As the use of images from actual surgeries was beyond present resources, cadaveric dissection of pig heads were used to generate images of the outcome of a surgical action. The use of cadavers however meant that any bleeding resulting from the skin being incised could not be simulated, as cadaveric skin has no blood supply. Nevertheless, as any bleeding will normally be cauterized and irrigated, the visual appearance should not be too dissimilar and the technique developed may potentially be applicable during actual surgery. Similar techniques may not necessarily extend to humans in a clinical environment. As such, clinical testing under more realistic conditions are necessary before this system can be deployed in an operation theatre.

While this work has shown promising results on the use of machine vision for inspecting the outcome of the three surgical tasks under consideration, there are several challenges regarding its use for this purpose. The first is the inherent natural variation in appearance of biological matter. A case in point is the inspection of incisions based on the presence of an incision line. Because of the large variations in appearance of the incision line, the reliance on a single feature was unable to capture all the differences between a complete and incomplete incision. The second is misclassification. False positives in the detection are more of an issue than false negatives. Erroneously classifying an incomplete surgical task as complete will result in the subsequent task being performed. When performing classification, the emphasis should therefore be on reducing the number of false positives, rather that achieving the best overall misclassification rate. The use of machine vision should therefore not be viewed in isolation. Indeed, the aim is to use machine vision to augment/complement other sensory information, although it is not inconceivable that the state of the art in machine vision can advance to a stage where it is able to emulate the visual recognition ability and interpretation process of a human expert.

References

[1] Glauser, D., Fankhauser, H., Epitaux, M., Hefti, J.L., Jaccottet, A.: Neurosurgical robot Minerva: first results and current developments. J. Image Guid. Surg. 1, 266-272 (1995).
[2] A. Blum et al., "Digital Image Analysis for Diagnosis of Cutaneous Melanoma. Development of a Highly Effective Computer Algorithm Based on Analysis of 837 Melanocytic Lesions," Br. J. Dermatol., vol. 151, pp. 1029-1038, 2004.
[3] F. Ercal et al., "Neural Network Diagnosis of Malignant Melanoma from Color Images," IEEE Trans. Biomed. Eng., vol. 41, pp. 837-845, 2002.
[4] H. Voigt and R. Classen, "Computer Vision and Digital Imaging Technology in Melanoma Detection," Seminars in Oncology, vol. 29, pp. 308-327, 2002.
[5] S. Narkilahti et al., "Monitoring and Analysis of Dynamic Growth of Human Embryonic Stem Cells: Comparison of Automated Instrumentation and Conventional Culturing Methods," Biomed. Eng., vol. 6, pp. 6-11, 2007.
[6] B. P. L. Lo et al., "Episode Classification for the Analysis of Tissue/Instrument Interaction with Multiple Visual Cues," Lect. Notes Comput. Sc., vol. 2879, pp. 230-237, 2003.
[7] N. Padoy et al., "On-Line Recognition of Surgical Activity for Monitoring in the Operating Room," in Proc. 20th Conf. Innovative Applications of Artificial Intell, Chicago, Illinois, 2008, pp. 1718-1724.
[8] W. O. Sack et al., Essentials of Pig Anatomy. Ithaca, New York: Veterinary Textbooks, 1982.
[9] G. M. Kaiser and N. R. Fruhauf, "Method of Intracranial Pressure Monitoring and Cerebrospinal Fluid Sampling in Swine," Lab. Anim., vol. 41, pp. 80-85, 2007.
[10] J. F. M. Manschot and A. J. M. Brakkee, "The Measurement and Modeling of the Mechanical Properties of Human Skin *in vivo*-I. The Measurement," J. Biomech., vol. 19, pp. 511-515, 1986.
[11] M. Mahy et al., "Evaluation of Uniform Color Spaces Developed After the Adoption of CIELAB and CIELUV," Color Res. Appl., vol. 19, pp. 105-121, 1994.
[12] R. M. Haralick et al., "Textural features for image classification," IEEE Trans. Syst., Man, Cybern., vol. 3, pp. 610-621, 1973.