World Academy of Science, Engineering and Technology
International Journal of Biomedical and Biological Engineering
Vol:4, No:9, 2010

Machine Vision for the Inspection of Surgical Tasks: Applications to Robotic Surgery Systems

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

***Abstract*—**The use of machine vision to inspect the outcome of surgical tasks is investigated, with the aim of incorporating this approach in robotic surgery systems. Machine vision is a non-contact form of inspection i.e. no part of the vision system is in direct contact with the patient, and is therefore well suited for surgery where sterility is an important consideration,. As a proof-of-concept, three primary surgical tasks for a common neurosurgical procedure were inspected using machine vision. Experiments were performed on cadaveric pig heads to simulate the two possible outcomes i.e. satisfactory or unsatisfactory, for tasks involved in making a burr hole, namely incision, retraction, and drilling. We identify low level image features to distinguish the two outcomes, as well as report on results that validate our proposed approach. The potential of using machine vision in a surgical environment, and the challenges that must be addressed, are identified and discussed.

***Keywords*—**Visual inspection, machine vision, robotic surgery.

I. Introduction

ROBOTIC surgery systems are capable of performing surgery autonomously under the direction and supervision of a surgeon, who intervenes in the event of a mistake. The MINERVA system [1], developed by the Swiss Federal Institute of Technology, is a robotic neurosurgery system capable of autonomously performing tasks such as incision and retraction of the scalp, and drilling of a burr hole through the cranium [2]. A burr hole is a small hole created in the cranium to provide access to the brain for surgery, and is a common neurosurgical procedure. Clinical trials of the system involving 13 patients have been reported [3, 4]. In the MINERVA system, force and torque characteristics were used to determine surgical task completion [2]. For example, in drilling of a burr hole, the force and torque profile of the drill is used to stop the drill bit after it has penetrated the cranium.

However, relying on these measurements alone is analogous to performing surgery blindly. The ability to mimic and embody a surgeon's visual sensory and decision making capability has the potential to improve current practice in robotic surgery. For example, machine vision could be used for visual inspection of the burr hole to validate successful drilling. Incorporating visual cues improves the reliability and robustness of robotic surgery systems by introducing redundancy in its sensor space, which traditionally only employs force, torque and positional measurements. As a step towards this goal, the use of machine vision to inspect the outcome of surgical tasks was investigated.

Several authors have investigated the use of vision in the operating theatre [5, 6]. Lo et. al. [5] developed a framework for the classification of surgical episodes in minimally invasive surgeries, in video sequences, using multiple visual cues. Specifically, their work was concerned with the detection of surgical tasks, such as the interaction of surgical instruments with soft tissue, retraction, cauterisation, and suturing. Padoy et. al. [6] developed an approach to recognize a subset of surgical tasks performed by the surgeon during laparoscopic surgery using visual cues. The aim of this study is to investigate a machine vision based approach for the inspection of the outcome of surgical tasks. Using machine vision in this context has not been reported. The use of machine vision, for greater situational awareness and confidence, has been the goal of our research.

Admittedly, the surgical environment can be highly unstructured, in contrast to the more controlled conditions typically found where machine vision is applied. However, because the surgical tasks being inspected are intended to be performed robotically, and their context is known, visual inspection is feasible. To illustrate the concept, practical examples of machine vision to inspect three primary surgical tasks involved in making burr hole, namely incision, retraction of an incision, and drilling of a burr hole, were used. The ability to inspect these tasks is an important functionality, from an operational and safety point of view for creating a burr hole robotically, as each of these tasks can only be performed upon successful completion of the previous task. For example, the scalp has to be properly incised prior to retraction, the incision properly retracted prior to drilling, and the cranium properly drilled prior to accessing the brain through the burr hole. The first illustrative example is the inspection of skin incisions.

M. Ovinis is a PhD scholar in the Department of Mechanical and Manufacturing Engineering, Loughborough University, LE11 3TU, Loughborough, UK (phone: +441509227579; fax: +441509227648; email: M.Ovinis2@lboro.ac.uk).

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. METHODOLOGY

A. Experimental setup

Experiments were performed using cadaveric pig heads to simulate the two possible outcomes (satisfactory and unsatisfactory) of three primary surgical actions involved in making a burr hole. Practical and ethical considerations prevent us from experimenting on humans or live animals. Thus, animal cadavers were used as a substitute. The selection of an animal cadaver should as far as possible be representative of human anatomy. With the exception of primates, the anatomy of the temporoparietal region (site of the burr hole) of pigs most closely resembles that of a human [7].

The surgical procedure for creating a burr hole in a pig as described by Kaiser and Fruhauf [8] was adopted with modifications. Tension was applied parallel to the plane of the skin to simulate parting of the skin (to facilitate placement of retractors after the skin is incised). Two square tabs (29 x 29 mm) were attached to the skin with cyanoacrylate adhesive [9]. The tabs were then pulled apart by two 600 gram weights using a pulley and weight system (Fig. 1). The distance between the tabs was 29 mm.

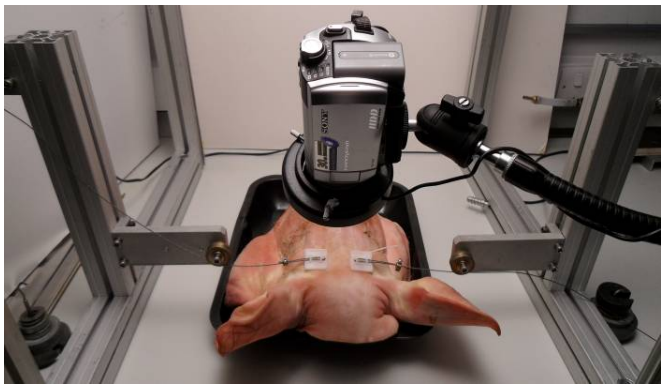


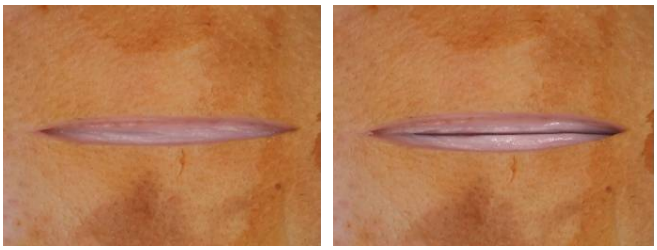
Fig. 1 (a) Experimental setup

The frontoparietal region was chosen as the site of the burr hole, instead of the temporoparietal region, because the underlying structure of the frontoparietal region allows for easier access to the brain. A midline incision, approximately 3 cm long, centred over the site of the burr hole, was made with a #10 scalpel blade, at different depths up to the pericranium. The incision was subsequently retracted at varying degrees using a self retaining retractor to expose the cranium, until it was wide enough to accommodate a burr hole. A burr hole 10 mm in diameter was then drilled in the cranium using a cordless drill at high speed. The characteristics of each surgical task were investigated to determine appropriate visual cues to distinguish the two outcomes. Because of the variability associated with biological subjects, fifty experiments were performed to give more statistically reliable data.

A CCD video camera, with a white LED ring-lighting unit mounted around the camera lens (Fig. 1), was used to obtain images of the surgical tasks. Used within its working distance, ring lightning can reduce the effects of shadows, by providing an even distribution of lighting in the field of view. The ring light had an adjustable intensity and a four section quadrant control. The white balance setting of the camera was manually adjusted to account for 6400K colour temperature of the white LED.

B. Characterisation of an incision

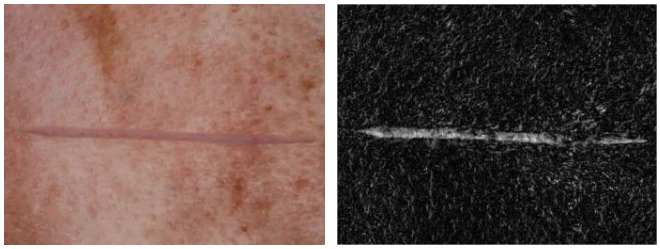
When surgeons incise the scalp, the scalpel blade is pushed into the skin, through the layers of the scalp, until it has touched cranium. A satisfactory incision should be both sufficiently long and deep i.e. an incision up to the pericranium along the length of the incision (Fig. 2b).



(a) (b)

Fig. 2 (a) Unsatisfactory incision (b) satisfactory incision

The shape properties and visual appearance of the two incision types were investigated. To determine the shape properties of the incision, it was first necessary to segment the incision in an image. Due to the variation in the appearance of skin, and because the edges of the incisions may have low contrast in an image (Fig. 3a); segmentation of the incision is not trivial. Segmentation of the difference image (obtained by subtracting an image before and after an incision is made) was attempted using region-based and edge-based segmentation. Because the underlying skin layer may be similar in appearance to the uppermost skin layer, compounded by the fact that the skin deforms when incised, the incision is ill-defined in the difference image (Fig. 3b).



(a) (b)

Fig. 3(a) Incision (b) Difference image on incision in (a)

Thresholding an intensity image of the incision, using a region-based local adaptive thresholding technique [10], and subsequently performing morphological operations gave better results (Fig. 4).

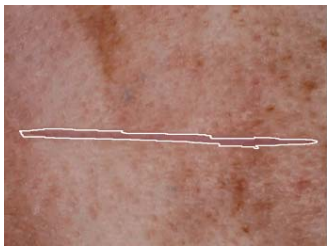
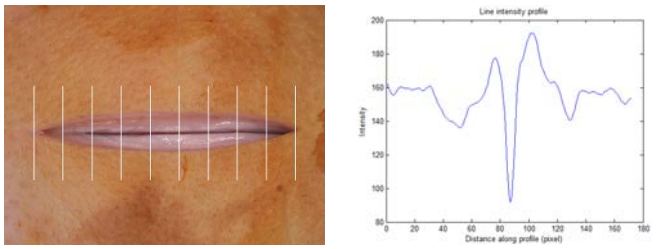


Fig. 4 Segmented incision using local adaptive thresholding

However, local adaptive thresholding is sensitive to threshold level and neighbourhood/filter size. As such, consistent results could not be obtained over a wide range of images. Difficulties in segmenting the incision notwithstanding, the shape properties of manually segmented incisions such as regional and boundary descriptors e.g. statistical moments, invariant moments, eccentricity, etc, were determined.

The visual appearance of the incision was examined next. The use of cadavers meant that any bleeding as a result of the skin being incised could not be simulated, as cadaveric skin has no blood supply. However, as any bleeding will normally be cauterised and irrigated, the techniques developed may potentially be useful during actual surgery. It was observed that satisfactory incisions exhibit a characteristic dark line, made as the scalpel blade incises the skin up to the pericranium, which is not present in unsatisfactory incisions. As such, the presence of a dark line along the length on an incision can be used to determine a satisfactory incision.

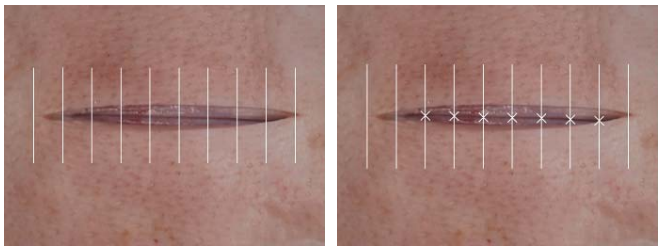
To detect the presence of this line, the line intensity profile of line segments along the midline of an incision was determined (Fig. 5). The pixel corresponding to an incision line is the global minimum of the line intensity profile. This method does not require the edges of an incision, which are difficult to segment, to be known.



(a) (b)

Fig. 5 (a) Line segment across incision (b) corresponding line intensity profile

The low contrast of the incision line in some incisions makes its detection difficult (Fig. 6). To enhance the contrast of the incision line, a contrast limited adaptive histogram equalisation (CLAHE) algorithm [11] was used.



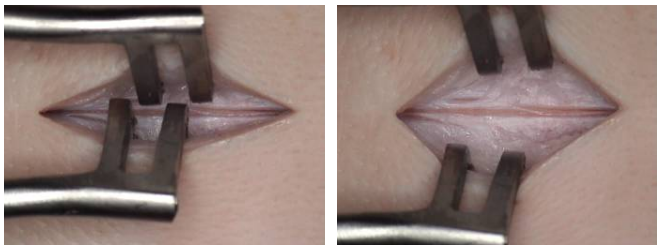
(a) (b)

Fig. 6 Output of incision line detection algorithm for (a) original image (b) contrast enhanced image

Assuming the camera is calibrated, the length of the incision would correspond to the distance between the first and last line segment containing a pixel that corresponds to the incision line. The resolution of the incision length determined by this method is dependent of the spacing between the line segments. The accuracy in determining the incision length is dependent on the accuracy of the calibration between the camera and task space, as well as the viewing angle of the camera, which ideally should be normal to the plane of view.

C. Characterisation of a retracted incision

Retraction of an incision (Fig. 7) is considered satisfactory if the incision is able to accommodate a burr hole. This may be indirectly determined from the separation distance of the prongs of the retractors. A potential problem with this approach is that the retractors may have opened without the incision being retracted i.e. the prongs of the retractor may not have engaged the edges of the incision, leading to the erroneous assumption that the edges of the incision are sufficiently separated. The approach adopted instead was to determine the amount of free space within the retracted incision by fitting the largest circle within the region bounded by the edges of the incision and prongs of the retractor. An incision is deemed to have been sufficiently retracted if the circle is larger than the prospective burr hole.

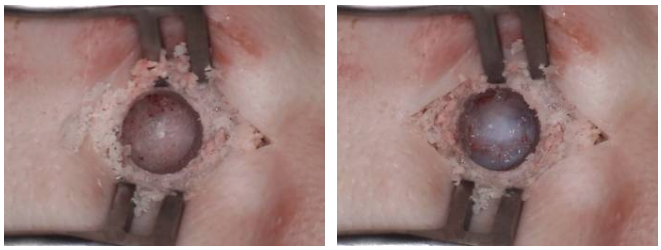


(a) (b)

Fig. 7 (a) Unsatisfactory retraction (b) Satisfactory retraction

D. Characterisation of a burr hole

A satisfactory outcome is one where the underlying dura mater i.e. the outermost part of the brain, is visible (Fig. 8b).

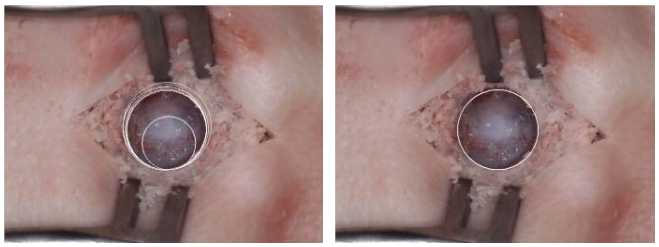


(a) (b)

Fig. 8 (a) Unsatisfactory/partial burr hole (b) Satisfactory/complete burr hole (dura mater is visible)

To detect the presence of a burr hole in an image, a generalised Hough transform to detect circles was used. The corresponding accumulator cell is a triple combination in 3D parameter space consisting of the circle centre coordinates and radius (range of radius values specified). Peak accumulator cell values represent a potential burr hole. A circle is disregarded if the ratio of pixels to perimeter is less than 10% of the total pixel number of the whole image. 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 in our images.

For each circle with a given radius, a minimum ratio of the number of accumulator cell count i.e. ratio of detected edge pixels to the number of pixels along the perimeter of the circle, was specified so that the algorithm was not biased towards larger circles, which may have a higher accumulator count over smaller circles. While the circle with the highest ratio of pixel count to the number of pixels along the perimeter of a circle usually corresponds to the actual burr hole, this is not always the case. This is because the edge of a burr hole is rarely a perfect circle (this is compounded by imperfect segmentation and distortions due to camera viewing angle), leading to the detection of multiple circles (Fig. 9a). By smoothing the accumulator array, we obtain a 'principle' circle (Fig. 9b). The principal circle was found to correspond well to the boundaries of a manually segmented burr hole. Once isolated, the difference in colour and texture of the two types of burr hole was investigated.



(a) (b)

Fig. 9 (a) Multiple circles detected (b) 'Principal' circle

III. Result and Discussion

A. Incision

A linear discriminant analysis was performed to classify 122 incision images consisting of 56 images of satisfactory incisions and 66 images of unsatisfactory incisions. The aspect ratio of the incision, defined as the ratio of the major axis and minor axis of an ellipse that has the same normalised second central moments as the incision, gave the best classification results. However, the misclassification error rate was 26.4%. Fig. 10 illustrates the aspect ratio for the two types of incision.

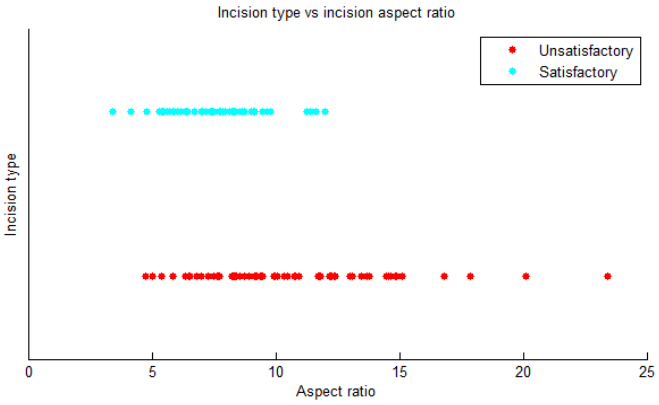
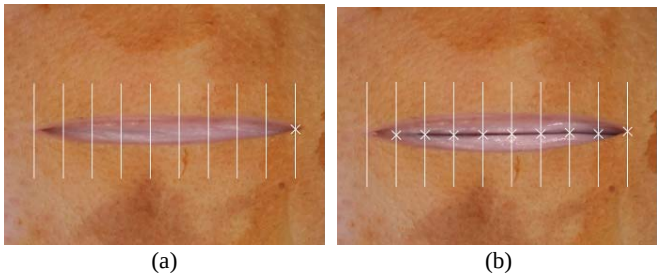


Fig. 10 Incision type versus incision aspect ratio

There is significant overlap in the feature space of the two classes of incisions. In general, it was found that shape properties were not discriminatory enough to distinguish the two types of incision. The in-class variation in shape properties is larger than the between class variation in shape properties i.e. two satisfactory incisions can have shape properties that are less alike than a satisfactory and unsatisfactory incision.

Fig. 11 illustrates the output of the incision line detection method for an unsatisfactory and satisfactory incision. The crosses represent the position along the line intensity profile that corresponds to the perceived position of the incision line.



(a) (b)

Fig. 11 Perceived incision line for (a) unsatisfactory incision (b) satisfactory incision

Classification of an incision using the line detection algorithm performed better, with a classification rate of 89.4%. Incisions were classified based on the mean distance of pixels corresponding to an incision line to a least squares fit

of a line to these pixels. Satisfactory incisions are incisions with a mean distance of less than half the line segment separation distance. Misclassification was mainly due to two reasons. The first is pixels being incorrectly identified as the incision line i.e. the global minimum of the line intensity profile does not always correspond to the incision line. The second is under detection of the incision line due to low contrast, even with contrast enhancement.

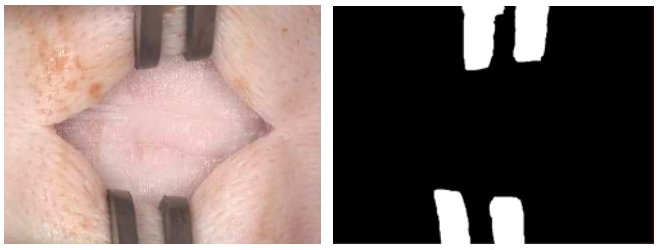
B. Retraction

A Canny edge detector was used to detect the edges of the incision. However, weak edges cannot be detected because of the sometimes low contrast at the boundaries of an incision. To improve segmentation of the incision, sections of the ring light were turned off to create shadows to facilitate detection of the edges of the incision (Fig. 12).

(a) (b) (c)

Fig. 12 (a) Left section of ring light turned off, (b) Right section of right light turned off (c) Detected edges

The next step was to segment the retractors. The retractors are made from surgical stainless steel and have a dull grey colour that is distinct in the operative site. The difference in colour of the retractor and the background was used as a basis to segment the retractors. Colour based segmentation [12] using a k-means clustering algorithm was used, with the initial cluster centroid positions or 'seeds' specified. The 'seeds' are the local maxima (determined using a hill-climbing optimization technique) of a three-dimensional L*a*b* colour space histogram of the image. Fig. 13 shows the segmentation results.

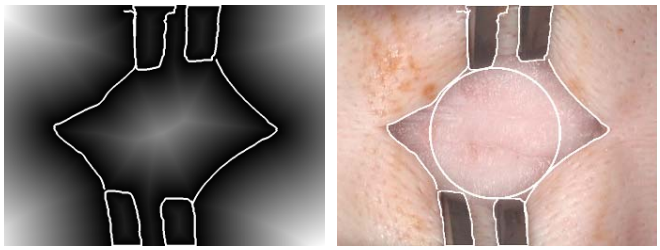


(a) (b)

Fig. 13 (a) Original colour image (b) Segmented image

Colour-based segmentation has successfully segmented the retractors by first dividing the colour image to regions of the most dominant colour. A representative colour value for the retractors was obtained by averaging the colour value of the pixels for a sample region of the retractor. This colour value was subsequently used to select the region whose average pixel colour value is closest to the representative colour value for the retractors.

Once the edges of the incision and retractors were detected, a distance map based on the distance transform of all the pixels within the incision, to the edges of the prongs of the retractors and edges of the incision was determined (Fig. 14a). The pixel with the maximum distance within these edges is the centre of the circle. The radius of the circle is the distance of this pixel to the nearest edge pixel of the prongs of the retractors or edge pixel of the incision (Fig. 14b). Assuming the calibration between the task space and camera frame is known, the diameter of the circle can be found. If the diameter of the circle is equal or larger than the diameter of the prospective burr hole (within a specified tolerance), the incision is deemed to be sufficiently retracted.



(a) (b)

Fig. 14 (a) A distance map overlaid with the edges of the incision and retractor (b) Circle whose centre corresponds to the pixel with the maximum distance within the incision in the distance map and whose radius corresponds to the distance of that pixel with the nearest pixel along the edges of the incision and retractor.

C. Burr hole

The visual inspection of the burr hole was formulated as a classification problem. The difference in colour between the cranium and dura mater was used to distinguish between a partial and complete burr hole. Each burr hole was represented by the median pixel colour value of all its pixels, in three colour spaces, RGB, L*a*b* and HSI. The images in the dataset were labelled for ground-truth and a leave-one-out cross validation using linear discriminant analysis (LDA) was performed to classify the images. Since the classification is based on statistics of the sample, classification results will be more reliable the larger the sample population. A random subset of 108 images was used (54 images of complete burr holes and 54 images of partial burr holes). Each image was used as training data with the remaining one being evaluated. The mean and maximum classification error was obtained by repeating the training and testing of the images ten times. Table 1 summarises the classification error when using the respective colour spaces.

TABLE I
COLOUR LINEAR DISCRIMINANT ANALYSIS LEAVE ONE OUT CLASSIFICATION ERROR RATE

Colour space	L*a*b*	HSI	RGB
Median (average)	2.4	2.6	2.4
(maximum)	(3.7)	(4.6)	(3.7)

The average misclassification rate in classifying burr holes based on colour is 2.4%. The classification sensitivity and specificity are 96% and 98% respectively. Although alternative classifiers might give better classification results, because of the relatively small dataset and the use of a single feature i.e. colour, the emphasis was on the strength of the features rather than strength of the classifiers. Classification rates were comparable in all colour spaces. The advantage of using the L*a*b* and HSI colour space over RGB colour space may only be apparent when there are changes in illumination, as changes in lighting will affect colour information. However, because the comparison is made for colour images without varying the illumination, it cannot be said that classification in L*a*b* or HSI space will perform better than classification in RGB space. Fig. 15 shows a scatter plot of the median pixel colour value of partial burr holes and complete burr holes in L*a*b* colour space, and the decision boundary.

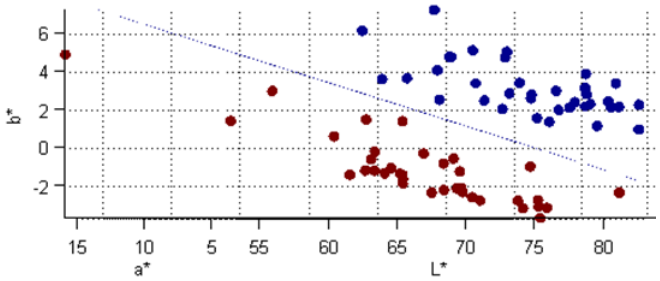


Fig. 15 Scatter plot of median pixel colour value of partial burr holes (black) and complete burr holes (light grey)

Fig. 16 illustrates all the images that the classifier could not discriminate based on colour alone.

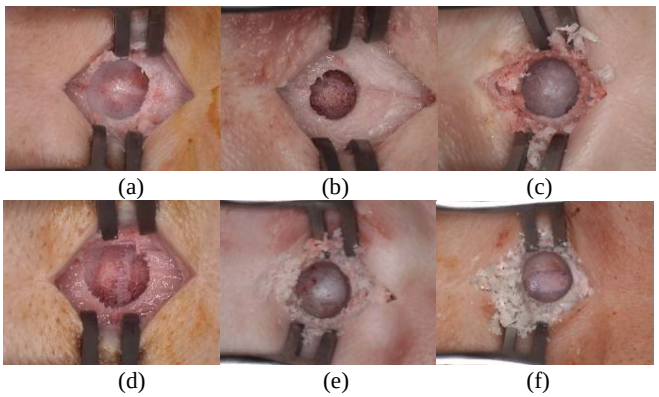


Fig. 16 Misclassified images

Fig. 16 a–e are false positives i.e. partial burr hole incorrectly classified as a complete burr hole, while Fig. 16 (f) is a false negative i.e. complete burr hole incorrectly classified as an incomplete burr hole. Textural information was considered to improve classification performance. Haralick's 14 coefficients [13] for the gray level co-occurrence matrix and statistical features for texture classification [14] were computed and used as a basis for classification. Classification was performed using all permutations of the different texture measures. The best classification results were obtained using the second moment/variance of the gray level image histogram of the burr hole, with a misclassification error rate of 11.9%. Combining textural information with colour information did not improve classification results.

Although this work has shown promising results on the use of machine vision for inspection of the outcome of the three surgical tasks, there are several challenges regarding the use of machine vision 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 satisfactory and unsatisfactory incision. The second is misclassification. False positives in the detection are more of an issue than false negatives. Erroneously classifying an unsatisfactory surgical task as satisfactory will result in the subsequent task being performed. The use of machine vision should therefore not be viewed in isolation. 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. This work represents the first step in this direction.

IV. Conclusion

In this paper, the use of machine vision to inspect surgical tasks was investigated. To the authors' knowledge, the use of vision in this context has never been considered before. Encouraging results were obtained for the surgical tasks under consideration, which has been demonstrated by our experiments. For example, the classification rate for incisions was 89.4% while classification rate for burr holes was 97%. We conclude that our findings are encouraging and suggestive for the potential use of machine vision for a more comprehensive approach in robotic surgery.

References

[1] C. W. Burckhardt, P. Flury and D. Glauser, "Stereotactic brain surgery," Engineering in Medicine and Biology Magazine, IEEE, vol. 14, pp. 314-317, 1995.

[2] N. Villotte, D. Glauser, P. Flury and C. W. Burckhardt, "Conception of stereotactic instruments for the neurosurgical robot minerva," in Engineering in Medicine and Biology Society, Vol.14. Proceedings of the Annual International Conference of the IEEE, 1992, pp. 1089-1090.

[3] H. Fankhauser, D. Glauser, P. Flury, Y. Piguet, M. Epitaux, J. Favre and R. A. Meuli, "Robot for CT-guided stereotactic neurosurgery," Stereotact. Funct. Neurosurg., vol. 63, pp. 93-98, 1994.

[4] D. Glauser, H. Fankhauser, M. Epitaux, J. L. Hefti and A. Jaccottet, "Neurosurgical robot Minerva: first results and current developments," J. Image Guid. Surg., vol. 1, pp. 266-272, 1995.

[5] B. P. L. Lo, A. Darzi and G. Z. Yang, "Episode Classification for the Analysis of Tissue/Instrument Interaction with Multiple Visual Cues," Medical Image Computing and Computer-Assisted Intervention: MICCAI .International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 230-237, 2003.

[6] N. Padoy, T. Blum, H. Feussner, M. O. Berger and N. Navab, "On-line recognition of surgical activity for monitoring in the operating room," in

Proceedings of 20th Conference on Innovative Applications of Artificial Intelligence (IAAI), 2008.
[7] R. A. Rival, O. M. Antonyshyn, J. H. Phillips and C. Y. Pang, "Temporal fascial periosteal and musculoperiosteal flaps in the pig: Design and blood flow inspection," J. Craniofac. Surg., vol. 6, pp. 466, 1995.
[8] G. M. Kaiser and N. R. Fruhauf, "Method of intracranial pressure monitoring and cerebrospinal fluid sampling in swine," Laboratory Animals(London), vol. 41, pp. 80-85, 2007.
[9] J. F. M. Manschot and A. J. M. Brakkee, "The measurement and modelling of the mechanical properties of human skin in vivo--I. The measurement," J. Biomech., vol. 19, pp. 511-515, 1986.
[10] P. Wellner, "Adaptive thresholding for the DigitalDesk," Xerox, EPC1993-110, 1993.
[11] K. Zuiderveld, "Contrast Limited Adaptive Histograph Equalization." in Graphic Gems IV. San Diego: Academic Press Professional, 1994, pp. 474-485.
[12] T. Ohashi, Z. Aghbari and A. Makinouchi, "Hill-climbing algorithm for efficient color-based image segmentation." in IASTED International Conference on Signal Processing, Pattern Recognition, and Applications, 2003, pp. 17-22.
[13] R. M. Haralick, K. Shanmugam, and I. Dinstein, “Textural Features of Image Classification,” IEEE Transactions on Systems, Man and Cybernetics, vol. SMC-3, no. 6, Nov. 1973.
[14] R. C. Gonzalez, & R. E. Woods, Digital Image Processing, Pearson Prentice Hall, pp. 828-842.