

PAPER • OPEN ACCESS

Vision Algorithms for Sensing Soft Robots

To cite this article: Victoria Oguntosin *et al* 2019 *J. Phys.: Conf. Ser.* **1378** 032102

View the [article online](#) for updates and enhancements.

You may also like

- [Aqueous Processing of \$\text{LiNi}_{0.4}\text{Mn}_{0.3}\text{Co}_{0.2}\text{O}_2\$ Composite Cathodes for Lithium-Ion Batteries](#)
Jianlin Li, Debasish Mohanty, Claus Daniel et al.
- [\(Invited\) Enabling Aqueous Processing of High-Energy and High-Power Electrodes for Lithium Ion Batteries – Issues and Mitigation Strategies](#)
Markus Börner, Lukas Ibing, Elizaveta Kessler et al.
- [Oxygen Control Challenge for Advanced Wet Processing](#)
Yukifumi Yoshida, Masayuki Otsuji, H. Takahashi et al.

Vision Algorithms for Sensing Soft Robots

Victoria Oguntosin¹, Ayoola Akindele, Olaitan Alashiri

¹Department of Electrical & Information Engineering, Covenant University, Ota, Ogun State
E-mail: victoria.oguntosin@covenantuniversity.edu.ng

Abstract.

The aim of this paper is the presentation and verification of computer vision algorithms in order to measure the geometric parameters of soft robots. The materials from which soft robots are made from possess large deformations. Embedded sensors or visual processing algorithms are often used to obtain measurement performance data from these robots. Integration of embedded sensors with soft robots can be cumbersome and expensive, also limiting the performance of a soft actuator. In this paper, implemented visual processing algorithms (thresholding, SAD, SSD and ZNCC) to measure performance data such as angle of motion, degree of bending, radius of curvature in real-time implemented with OpenCV libraries and Webcam is described. Soft RGB colour markers were also produced and firmly glued into the body of the soft robot with no hindrance to movement. Some concepts of visual processing applied include colour tracking, template matching and camera calibration. The execution of vision based motion control to a variety of soft actuators such as bending and wedge-shaped soft actuators was described.

1. Introduction

Vision is a low cost and adequate way of sensing the world. Using a camera as a sensor is arguably one of the most sophisticated method of sensing an environment. A camera is a passive, exteroceptive sensor, it receives light signal from the environment and measures the amount of light reflected by an object.

Soft robots are made with deformable bodies and have a number of advantages such as flexibility, dexterity and inherent compliance. However, one downside to the extra flexibility of a soft robot is that this creates difficulty in accurately measuring position and movement paths. One of the problems in soft robotics is finding a suitable way of measuring the parameters and motion of these robots. A robust, standardized, stable measurement and control system is still a challenge. Although there is on-going work in developing standardized sensors, there is yet the need to adapt and integrate these sensors to a specified design at prototypes scale because soft roboticists still have to fabricate their soft robots from scratch. And if soft structures do not have reliable and standard sensors, then control would become very difficult. Our work focuses on addressing this problem by applying standard computer vision algorithms [1] in measuring the structure and motion of soft robots. These computer vision techniques are basic, readily available in standard computer vision libraries and therefore could easily become a unified, standardized method for sensing soft robots.

When it comes to robotics, we often borrow from nature and this is another classic example of how we borrow from nature. Our application of a vision sensor system in this work is to provide a low cost method in order to collect real time parameters of a rubber-based soft robot material by solely the use of vision as a sensor. The complete setup of our application is low cost; computational requirement is taken care of by the operating PC; the Webcam is powered via USB connection to the PC; the weight of and size the Webcam is such that it is portable and easy to handle.

Computer vision has been applied in an extensive variety of ways, some of which includes industrial applications for inspecting manufactured objects, surveillance and robotics. Computer vision techniques have been applied to measure the geometric parameters of a soft robot. For example, a colour intensity value of a soft robot that changed colour upon actuation was measured using a digital camera [2]. In [3, 4], curvature formula was derived and used in curvature calculation of a continuum robots similar in

¹ Corresponding author.



design to an elephant's trunk. Vision was applied to determine a soft robot segment curvature as it undergoes deformation to a curved shape in [5], with the use of this visual information; a control law was applied to achieve a desired curvature. Vision-based motion control systems have been used in robotic applications to measure and control plate positions of pneumatic powered actuators [6]; to rotate a humanoid robotic head [7]; to implement controlled movements of a flexible robot through accurate 3D shape reconstruction and spatial localisation in real time [8]; for real-time visual tracking [9]. Likewise, the human arm motion tracking through camera recording of colour markers attached to the arm itself has been carried out [10].

Visual tracking of objects in motion can be discriminatory or generative, while a discriminatory approach reduces the tracking to binary classification; generative approach makes use of a region of interest to locate in subsequent frame. Visual tracking relies on intensity and texture information. To determine the position of a marker, a template-based matching algorithm implemented using absolute or squared difference or normalised cross correlation or ridge regression can be applied [11]. Template matching is a robust method that excels with low textured objects but comes with increased computational cost and therefore not well adapted for real-time applications [12]. A type of matching cost that applied normalized cross correlation was developed in [13] to handle noise, colour variation, inconsistent shadow and reflection from object; to consider texture information in order to get reliable feature points. Zero normalized cross correlation is robust to lighting, compensates for local gain and offset changes between matched image patches with great precision and provides the ability to handle large radiometric differences in an input image [14]. Points of interest across two images were matched together [15] using ZNCC. This cost computation method was applied for dense stereo matching using frame-to-frame tracking in [16]. Though it is more conservative than other methods of matching such as sum of absolute or square differences in uniform regions, it is more accurate in areas where noise might be important in an image, e.g. textured areas.

A more popular and widely used method of determining the position of a particular tracked point is using only colour space information to reduce the tracking into binarised form using thresholding. It makes use of luminance and colour intensity information. Colour measurements vary considerably over an image sequence due to illuminance, shadows, shading and specularities variations [17]. These variations are normally eliminated with several preprocessing prior to object identification.

Other methods applied to track features with a camera can be linear [18] or Ridge Regression [19]. Ridge regression minimizes a cost function and it includes mean-square-error and Tikhonov regularisation term. This method was applied in [20] to track myography movements on the surface of the forearm. Some other feature point detection algorithms include SURF [21], SIFT [22], amongst others. Marker systems for object detection are template based markers or 2D fiducial digital marker-based such as ARTag [23]. They both provide precise and robust tracking methods. Digital marker systems are not employed in this work because of the simplicity our application as digital markers would require more computation and less suited for real time usage. Template based method was more suited due to easy integration and ability to produce markers of the same material with the soft robot. Commercially available vision tracking systems include OptiTrack (Corvallis, United States) and Vicon (Oxford, United Kingdom)

The contributions of this paper include the application of standard vision processing algorithms to determine the position of silicone based soft robots with trackers integrated into the body of the robots used to achieve little or no resistance to movement. We use a low cost system to implement this work. We described techniques for real time processing for computational expensive algorithms. We showed results of our system measuring parameters for a variety of soft structures such as bending and wedge-shaped actuators. Finally, we used results obtained from the visual measurement system to execute a two-stage control law that can be used in the movement of soft robot to attain a desired position.

2. Methodology

Visual Processing Algorithms for colour intensity and template-based tracking were used as a means of finding the soft robots position and current state.

2.1. Colour Intensity Tracking:

This method is aim at differentiating the specific colour marker from the background by reducing the scene to a binary image. Colour tracking can be achieved by using distinctive colour markers such as Red-Green-Blue (RGB) colour markers in the scene and performing thresholding to binarize the image—the image is separated into two pixel classes with a minimal variance. The pixel classes are black and white. White indicates the colours being tracked while black is everything else. The pre-processing and post-processing algorithms for thresholding method include:

1. **Filtering:** Filtering is applied to reduce noise by averaging each pixel of the original image over a odd-number square matrix window. Gaussian kernel is applied to make the averaging isotropic with decreasing weight away from the tracked point. Median operation through ranking can also be used to preserve detail during the filtering.
2. **RGB to HSV Conversion:** Hue Saturation Lightness (HSV) is an intuitive way of specifying colours rather than RGB. The tracking is carried out in the HSV colour space in order to separate lumia intensity from chroma intensity. In HSV colour space, hue selects the colour; saturation selects how deep and rich the colours are (fully de-saturated is white, fully saturated is pure colour); and value behaves like a dimmer. In this colour space, H and S are invariant for shadow-shading.
3. **Binarisation:** A logical operation is performed to separate the image into black and white. This logical operation is used in a vectorised form. The logical test is done between every element of the image input matrix and a vector matrix. The vector matrix specifies the range of HSL values of the red, green and blue colour trackers. The output matrix becomes a logical value (true or false).
4. **Erosion and dilation operation:** This operation is carried out on the binarised image to open and close holes with a square matrix morphological structuring element, S, sensitive to shape. Erosion and dilation are converse operations. For erosion, the output is positive if all pixels in S are 1s. For dilation, the output is positive if any pixels in S are 1s. The arrangement of erosion trailed by dilation is called opening, only shapes compatible with the structuring element are maintained while the rest shapes disappear. The arrangement of dilation trailed by erosion is called closing; holes within the shape are closed up.
5. **Find binary blobs:** To locate the geometry of tracked object, blobs have to be determined. A blob is defined as a group of contiguous pixels in a binary image, connected to each other which have the same colour. Since multiple colours are being tracked, the image has multiple blobs distinctive from each other by the RGB thresholding operation.
6. **Find centroid of blob:** After defining blob regions in the image, the position coordinate of the blobs are found by calculating the zeroth and first moments. The zeroth moment, m_{00} is the number of one (true) pixel in a region. This specifies the area of the tracked region, given as:

$$m_{00} = \sum_{(m,n) \in I} I[m, n] \quad (1)$$

Where m and n coordinates are the $x - y$ locations of pixels in the image and $I[m, n]$ specify the cost of a binary pixel at that precise location. The first moments, m_{10} and m_{01} are weighted averages of m and n coordinates, weighted by the colour of binary pixels and specified as:

$$m_{10} = \sum_{(m,n) \in I} uI[m, n] \quad , \quad m_{01} = \sum_{(m,n) \in I} vI[m, n] \quad (2)$$

The coordinate of the geometric centre or centroid of the tracked region, $[m_c; n_c]$ is given by:

$$m_c = \frac{m_{10}}{m_{00}}, \quad n_c = \frac{m_{01}}{m_{00}} \quad (3)$$

2.2. Template Matching:

Template matching is a generative approach to tracking as it concentrates on examining the regions in an image which are comparable to the tracked image [18]. Template matching calculates a correlation that represents the degree of resemblance or difference to a template patch at all positions of the image [24].

The location tracking of binarised colour markers, whilst efficient requires constant recalibration to account for pose variation. A robust method, though more computationally expensive, is using a template or pattern of pixels to find the location of tracked points. The implementation also requires the use of colour markers. This is a spatial operator approach in which the input image is compared with a template patch, The template is the image being looked for, the input window is a subset matrix of the input image that would have the same size with the template. In other words, a numeric similarity score between a particular input window of the image which matches the template is found. The tracked colour marker template is compared across every single location in the scene. To do achieve this comparison, an image similarity measure is needed.

Sum of absolute Differences (SAD) is an image similarity measure that takes the absolute difference between corresponding pixels of the 2 images being compared and sums it up. The similarity measure, s , has a value of zero if the both images are identical and a value greater than zero if dissimilar. In other words, the position where the minimum value of the output similarity matrix is located indicates the position of a perfect match. The equation is specified as:

$$SAD, s = \arg \min \sum_{(m,n) \in I} |I_1[m, n] - I_2[m, n]| \quad (4)$$

Another related method is the operation called Sum of Squared Differences (SSD), in the place of using absolute difference, the sum of squared difference is calculated as:

$$SSD, s = \arg \min \sum_{(m,n) \in I} (I_1[m, n] - I_2[m, n])^2 \quad (5)$$

Another measure called Zero Mean Normalized Cross Correlation (ZNCC), sums the difference divided by the square root of the squared product of the input and template window. The ZNCC score size varies from -1 to +1. +1 implies that the two image patterns are matching; -1 implies that both images are negative of each other. 0 translates to the two images being not well correlated. A characteristic value of 0.8 will be measured to be a realistic match with the template. Its equation is given as:

$$ZNCC, s = \arg \max \frac{\sum_{(m,n) \in I} I_1[m, n] - I_2[m, n]}{\sqrt{\sum_{(m,n) \in I} I_1^2[m, n] \cdot \sum_{(m,n) \in I} I_2^2[m, n]}} \quad (6)$$

If an object template is applied to another scene having the same object template but in a different position or slightly modified such as change in lighting conditions or camera position to make the tracked point bigger if it is closer to the camera or smaller if it is farther from the camera, template matching would not work well, hence the need for constant recalibration. Recalibration is easier using template matching as the tracked region is selected in online prior to tracking.

3. Implementation

The tracker markers were RGB coloured markers made with silicone rubber, similar the material with which the soft robot is made. This was manufactured by adding and mixing few drops of coloured paint

to a silicone liquid to produce individual red, blue and green soft colour trackers. Ecoflex 0030 silicone liquid is generally transparent and can be coloured as desired with pigment pastes. The pigment is easily blended into the silicone liquid compound during mixing. The soft colour markers were glued to the body of the soft robot as shown in Figure 1.

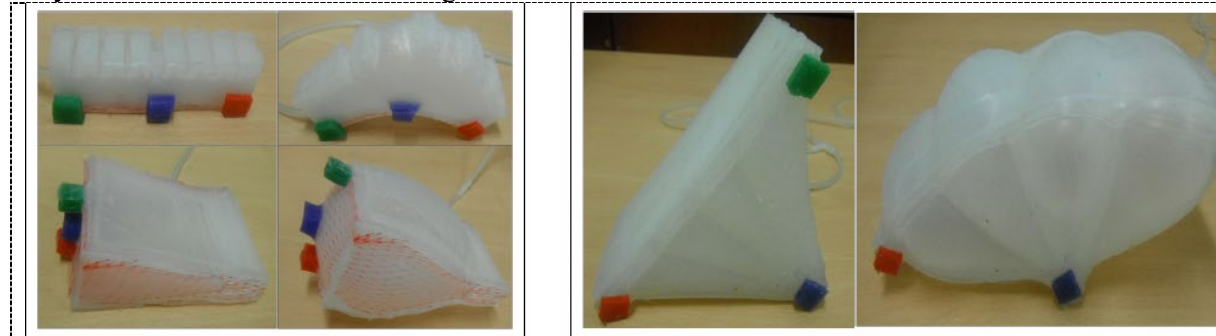


Figure 1. Soft colour markers bonded with the soft robots to measure parameters of various soft robots as they undergo movements through inflation and deformation.

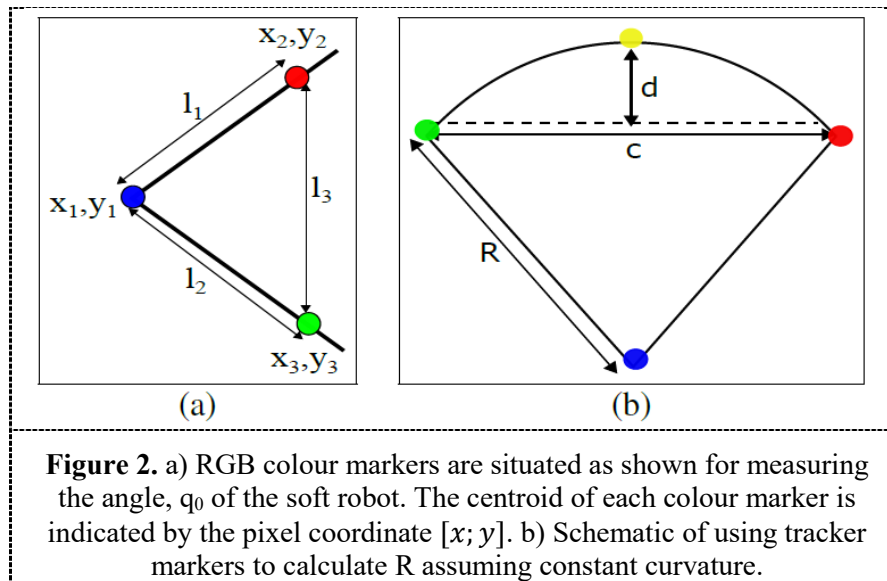
The visual processing was carried out on a Windows Computer having an Intel quad core i5-4460 processor having a processor speed of 3.2GHz. The setup employed is extremely simple; a standard Webcam views the scene. When the camera axis is orthogonal to the object plane, and at a fixed distance, the geometric setting of the camera relative to the object can be selected in order that pixel units can be converted to metric value. [25]. A perspective camera (like the human eyes) will always project a straight line to a straight line. It will project a conic section to a conic section, though not necessarily the same conic section, so a circle (a conic section) could be mapped to a different conic section such as an ellipse. Angles between lines are not preserved which means parallel lines may be projected as converging, and the vertex angles of a planar shape (triangle, square or more generally a polygon) may not be preserved.

The amount of light that is reflected off a surface depends not only on the colour of the object but also on its roughness, distance of light from the object and most importantly the colour of the light source and ambient light. The pixel value of a point in the scene can change even if the camera does not move. A slight change in illumination, perhaps a difference in sunlight or light source can change the exposure of that point in the scene. So calibration is important because of these highlighted factors are constantly changing. To accommodate for all these factors, recalibration is performed each time by selecting the group of pixels to be tracked, that way, ambient light levels is taken care of and other discrepancies. The template matching software incorporated a feature for real time selection of the tracked group of pixels.

3.1 Angle of motion

Three-colour tracking was carried out by the RGB colour markers to measure angle. The coordinate position of the tracked marker in camera view is defined by the centroid of each colour marker. Figure 2 illustrates the image camera view for measuring the angle which is calculated as:

$$q_0 = \cos^{-1} \frac{l_1^2 + l_2^2 - l_3^2}{2l_1l_2} \quad (7)$$



The lengths l_1 , l_2 and l_3 are measured as the distance, pixel units, between the center of the R and B; B and G; R and G colour markers respectively. In order to achieve better results while using computer vision to get angle measurements, a recalibration of the HSV values of the 3 colour markers was carried out at all instances the experiment was done in order to minimise the effect of change in lighting conditions of the room.

3.2 Radius of Curvature

The radius of curvature, R , of a curve at a point is defined as a measure of the circular arc radius which best approximates the curve at that point which is also the inverse of curvature. The radius of curvature from intersecting chords theorem is given as:

$$R = \frac{c^2 + 4d^2}{8d} \quad (8)$$

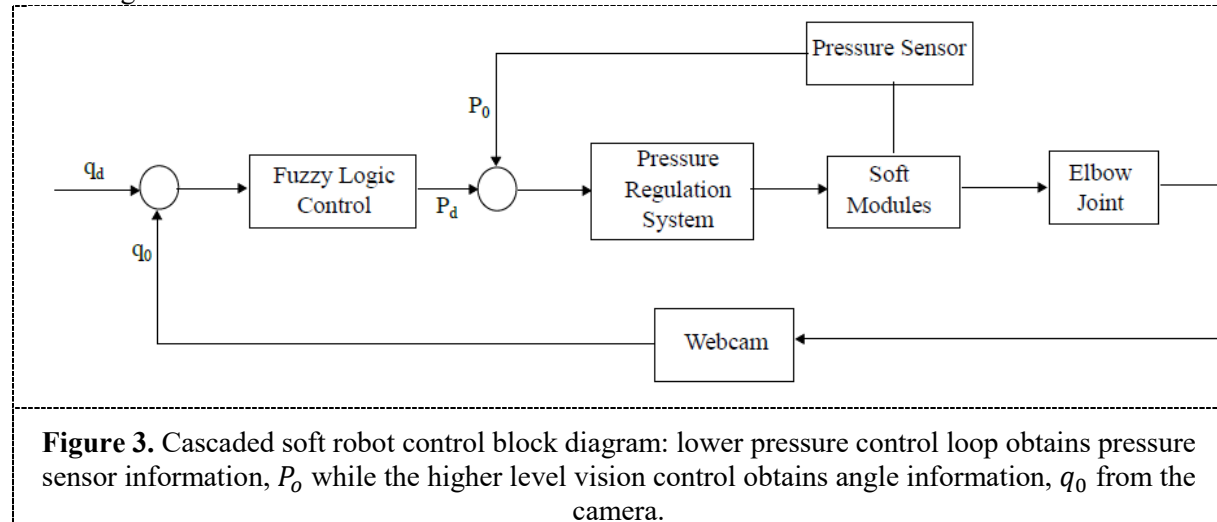
Where c is the arc's width, and d is its height, both measured from the visual processing in pixel units as shown in Figure 2b. The arc length, L , is then calculated as:

$$L = 2\pi R \frac{\theta}{360} = 2\pi R \frac{2 \sin^{-1}(\frac{c}{2R})}{360} \quad (9)$$

3.3 Control Scheme

The pressure and angle measurement system consists of 4 aspects: camera, computer for vision processing, air pressure controller and manufactured soft robot. The camera relays vision information, through visual processing. The air pressure controller consists of a pump, ON/OFF solenoid valves, pressure sensor and controller. Established serial communication through USB connection of a camera to a computer was used to get the measured angle q_0 from the computer to the controller which subsequently generates an applicable PWM signal that activates the air pressure controller system as air is fed into the soft robot. This cascaded controller intermittently obtains differences between the measured angle q_0 and desired angle q_d and directs the PWM signal which is directed into the air pressure control system for a resolution of the discrepancy. An error signal: $e = q_d - q_0$, is sent to the air pressure controller which in turn generates an output signal in proportion to the PWM signal

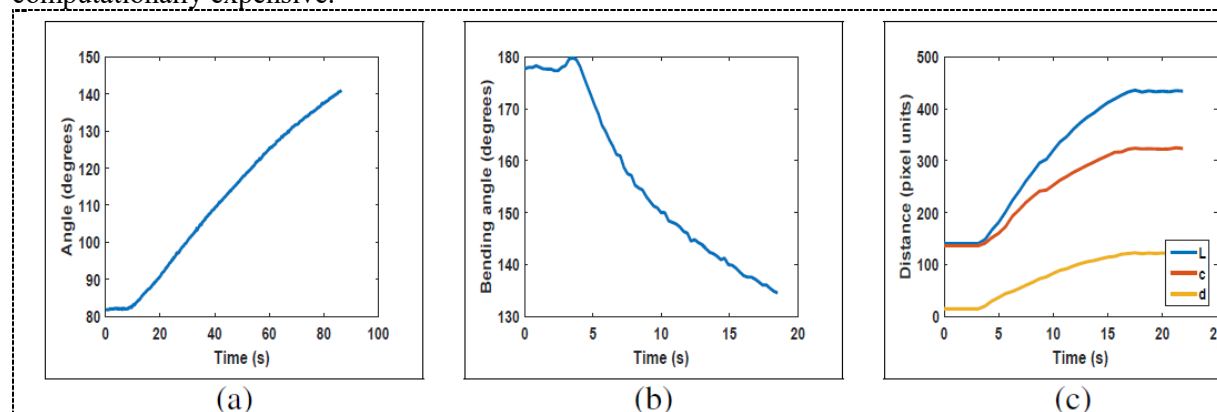
generated to control air pressure within the soft robot. Figure 3 illustrates the cascaded soft robot control block diagram.



4. Results and Discussion

We solely rely on the developed visual processing software for all manufactured soft robots which includes the bending robot, 1-segment triangle-shaped robot having an angle of 20° and the 80° 4-segment triangle actuator as shown in Figure 1. The soft colour markers which also formed the body of the robot were fixed to specify tracked points. Figure 4 shows graphs of the visual processing software measuring angle, bending distance and arc length as a function of time.

Both methods of colour tracking via thresholding and template matching were used. The template matching method provided more accuracy in the tracked point; real time calibration was possible using this method as the template was selected using mouse selection. Template matching was also more computationally expensive.



In general, template matching gave fewer errors and was more reliable but it is slower. The time to process each frame for three colour markers for template matching is 600ms while thresholding

algorithm was about 280ms. When frames were converted from RGB to grayscale; the speed of visual processing for template matching improved to a value between 100 – 120ms. This processing time measured was reduced because in RGB colour tracking, three values were being compared but in grayscale instead of three numbers, it is just a single number. This is because grayscale represents the image as a 1-Dimensional array as opposed to an RGB colour image which is represented as a 3-Dimensional array of pixel values. Another important factor affecting visual processing speed is the size of the image, a larger image of 640 by 480 slows down the frame rate compared with a 320 by 480 image.

The template matching performed better as room lighting conditions would have a lesser chance in affecting obtaining the correct location of tracked regions when compared with thresholding. Calibration is easier with template matching, no need of manual adjustment and recalibration of HSV values due to changes in the lighting conditions of the room. During experiments template was also found to be a more robust method. As a way of minimising errors while measurement data and control was carried out with the template matching code, at the start of the experiment, a test run was carried out to ensure that the correct position of the moving marker is tracked through the entire range of motion. The tracking template is then adjusted or the position of the soft colour marker.

Visual processing was used not only as a sensor but also as an aid to perform high level control of the soft robots. Based on the current positional (angle, bending distance or arc length) measurement, a proportional control was implemented which produced a control signal to be used as input to a low level pressure controller that inputted the desired pressure into the soft actuator.

Conclusion

We presented vision algorithms and concepts that were utilized in sensing soft actuator parameters in real-time. The other possibilities of utilizing computer vision for soft robots can be measuring the shape of deformation by placing markers along the curve. One of the advantages of using vision for our system is that it is cheap. Soft robots themselves are very cheap to manufacture, so using the low cost real time vision system will offer a suitable complement. In addition, due to the emerging nature of soft robotics and having no standardized sensors available, computer vision offers a valid alternative.

References

- [1] Corke, P. (2011). *Robotics, Vision and Control, Fundamental Algorithms in MATLAB*, 73(1), Springer Tracts in Advanced Robotics.
- [2] Gossweiler, G. R., Brown, C. L., Hewage, G. B., Sapiro-Gheiler, E., Trautman, W. J., Welshofer, G. W. & Craig, S. L. (2015). Mechanochemically active soft robots. *ACS Applied Material Interfaces*, 7(40), 22431–22435.
- [3] Hannan, M. W., & Walker, I. D. (2003). Kinematics and the implementation of an elephant's trunk manipulator and other continuum style robots. *Journal of Robotic Systems*, 20(2), 45–63.
- [4] Jones, B & Walker, I. (2006). Kinematics for multisection continuum robots. *IEEE Transactions on Robotics*, 22(1), 43–55.
- [5] Marchese, A. D., Komorowski, K., Onal, C. D., & Rus, D. (2014). Design and control of a soft and continuously deformable 2D robotic manipulation system. *IEEE International Conference on Robotics and Automation*, 2189–2196.
- [6] Hirai, S., Shimizu, K., & Kawamura, S. (2002). Vision-based motion control of pneumatic group actuators. *IEEE International Conference on Robotics and Automation (ICRA)*, 3(1), 2842–2847.
- [7] Visser, L.C., Carloni, R., & Stramigioli, S. (2009). *Vision based motion control for a humanoid head*. *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 5469–5474.

- [8] Vandini, A., Salerno, A., Payne, C.J., & Guang-Zhong, Y. (2014). Vision-based motion control of a flexible robot for surgical applications. *IEEE International Conference on Robotics and Automation (ICRA)*, 6205–6211.
- [9] Inaba, M., Nagasaka, K., Kanehiro, F., Kagami, S., & Inoue, H. (1996). Real-time vision-based control of swing motion by a human-form robot using the remote-brained approach. *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 1(1), 15–22.
- [10] Zanchettin, A.M., Bascetta, L., & Rocco, P. (2013). Achieving humanlike motion: Resolving redundancy for anthropomorphic industrial manipulators, *IEEE Robotics Automation Magazine*, 20(4), 131–138.
- [11] Patil, S., Nadar, J. S., Gada, J., Motghare, S., & Nair, S. S. (2013). Comparison of various stereo vision cost aggregation methods. *International Journal of Engineering and Innovative Technology*, 2(8), 222–226.
- [12] Hinterstoisser, S., Holzer, S., Cagniart, C., Ilic, S., Konolige, K., Navab, N., & Lepetit, V. (2011). Multimodal templates for real-time detection of texture-less objects in heavily cluttered scenes. *IEEE International Conference on Computer Vision (ICCV)*, 858–865.
- [13] Xiaoyong, S., Li, X., Qi, Z., & Jiaya, J. (2014). Multi-modal and multi-spectral registration for natural images, *Lecture Notes in Computer Science*, Springer International Publishing 8692 (1), 309–324.
- [14] Banks, J., & Corke, P. (2001). Quantitative evaluation of matching methods and validity measures for stereo vision. *The International Journal of Robotics Research*, 20 (7), 512–532.
- [15] Lhuillier, M., & Quan, L. (2002). Quasi-dense reconstruction from image sequence, *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, 2351 (1), 125–139.
- [16] Pradeep, V., Rhemann, C., Izadi, S., Zach, C., Bleyer, M., & Bathiche S. (2013). Monofusion: Real-time 3d reconstruction of small scenes with a single web camera. *IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 83–88.
- [17] Danelljan, M., Khan, F.S., Felsberg, M., & van de Weijer, J. (2014). Adaptive color attributes for real-time visual tracking. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1090–1097.
- [18] Wang, D., Lu, H., & Ming-Hsuan, Y. (2013). Least soft-threshold squares tracking. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2371–2378.
- [19] Hoerl, A. E., & Kennard, R.W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 121:55–67.
- [20] Nissler, C., Mouriki, N., Castellini, C., Belagiannis, V., & Navab, N. (2015). OMG: Introducing optical myography as a new human machine interface for hand amputees. *IEEE International Conference on Rehabilitation Robotics (ICORR)*, 937–942.
- [21] Bay, H., Tuytelaars, T., & Van Gool, L. (2006). SURF: Speeded up robust features, *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, 3951(1), 404–417.
- [22] Ke, Y., & Sukthankar R. (2004). PCA-SIFT: A more distinctive representation for local image descriptors, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2(1), 506–513.
- [23] Olson, E. (2011). APRILTAG: A robust and flexible visual fiducial system, *IEEE International Conference on Robotics and Automation (ICRA)*, 3400–3407.
- [24] Stefano, L. D., Mattoccia, S., & Tombari, F. (2005). ZNCC-based template matching using bounded partial correlation, *Pattern Recognition Letters*, 26(14), 2129 – 2134.
- [25] Treiber, M. A (2013). *Optimization for Computer Vision*, Springer.