34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

Application of scale-space descriptors for the reliable detection of keypoints for image registration in optical mapping studies in whole heart preparations

M. Rodriguez, A. Nygren, *Member, IEEE*

Abstract— **Data acquired using Optical Mapping (OM) studies are affected by motion artifacts due to the inherent contraction of the heart. Those artifacts can be reduced by registering the images obtained by the OM system or by the combination of approaches like physical restraint of the heart or ratiometry with image registration. Due to the lack of high contrast features most registration methods are not suitable for this application. This paper is focused on the utilization of scale space theory and local descriptors to enhance the detection of local features in OM images and to describe the movement of keypoints. This information can be used to determine a suitable set of transformations to perform the registration process.**

I. INTRODUCTION

Optical Mapping (OM) is used as an imaging modality based on the fluorescence property of voltage sensitive dyes that bind to the cell membrane. Cardiac OM provides information related to the spread of electrical activity in the heart. This is helpful for understanding electrophysiological dynamics at the whole heart level, which gives insights into arrhythmias including initiation and maintenance [1][2]. One disadvantage of this modality is the presence of gross motion artifacts in the recorded optical signals due to the inherent contraction of the heart during the experiments.

The artifacts appear as a distortion of the membrane voltage signal, or other parameters being studied that have been extracted from the change of the pixel intensity in the images through the movie. The video is created with the optical images from the same scene at different times. The contraction of the heart produces a movement of the pixels being observed and the signal extracted appears distorted since the information is collected from different but close regions in the heart. Some approaches for motion artifact removal or improvement are [3][4][5]: (1) mechanical constraint, (2) chemical uncoupling to block contraction, (3) image registration, and (4) dual wavelength ratiometry.

The use of chemical agents to prevent contractions tend to affect the electrophysiology [3] of the heart, affecting directly the information extracted from the signals. The fact that image registration can be applied post imaging, does not affect the physiological response of the heart, and can be utilized with conventional and dual wavelength OM, depicts some of the advantages of using this technique as a complementary method to attenuate motion artifacts in the optical signal.

Different image registration techniques have been applied for OM images, e.g., basic cross-correlation of regions of interest (ROI) [3], optimization of mutual information [4] maximization of cross correlation coefficient as a measure of similarity [5] and non-linear methods [6], most of these studies showed results of registration for a specific ROI inside the imaged heart, i.e, the studies do not describe the motion across the entire heart. The control points or regions being registered are chosen manually in most of the cases with whole heart preparations. Approaches include the selection of fixed small ROI in the frame sequence [3], the selection of one specific ROI to be registered [4], finally [5] using the entire frame but locating markers manually to elucidate the motion at some points inside the frame.

In this paper, we present an approach for the automatic selection of the control points to be used to calculate the set of transformations required to register the images. The methods used to track the control points' movement throughout the sequence of frames are also presented.

II. METHOD - IMAGE REGISTRATION AND SCALE SPACE THEORY

A. Optical System and heart preparation

This study was done in whole Langendorff perfused rat hearts, the preparation used Di-4-ANNEPPS as the voltage sensitive dye and the imaging data was obtained from a CA-D1-0128T Dalsa camera and a 250W quartz tungsten halogen light as the illumination source. The images obtained

The authors would like to acknowledge the support of NSERC.
M. Rodriguez and A. Nygren are with the Department of Electrical Engineering of the University of Calgary.
A Nygren is with the Centre for Bioengineering Research & Education, Schulich School of Engineering, University of Calgary. AB, Canada. E-mail: nygren@ucalgary.ca
*M. Rodriguez is being sponsored by COLCIENCIAS-LASPAU.

were 60x60 pixels with a 12 bits resolution at 950 frames/s [7], the spatial resolution is 250µm per pixel with a field of view15x15 mm.

B. Scale Space

Studies have shown that it is possible to remove motionartifacts in specific regions of OM images [3][4][5][6]. However, these results cannot be extrapolated to whole heart images since the error measures tend to increase with time, i.e., with the position of the image in the frame sequence. The reason for this is the lack of information of the individual motion for different regions of the heart.

Fluorescence images lack contrast and the nature of the images does not give rise to enough contrast features, required for most techniques to calculate motion. Scale space theory is a reliable technique to perform matching of control points inside an image of the same scene at different times.

The use of Scale-Space theory reveals information in the image that is invariant to some transformations like rotation, translation, and scale [8]. This theory uses kernels to detect information. The convolution with Gaussian functions (GF) of the images provides the scale σ, parameter of the GF, which acts as the transformation operator, this applied in cascade generates the kernel.

This paper presents two Scale-Space detectors, (1) Scale Space Feature Transform (SIFT) and (2) Speeded-Up Robust Features (SURF). These were implemented and modified to work with the low contrast low resolution fluorescence images.

C. SIFT

The implementation of this descriptor is based on the description made by Lowe in [9]. Once the Scale-Space pyramid has been created, the first step is to select the key or interest points in the image; this is achieved by detecting the scale space extrema, then an interpolation process is performed. Calculation of the descriptor is done for every keypoint and finally it is possible to find the matches in two scenes having the descriptor vectors for each image.

Scale-Space Pyramid: This is created by progressively convolving the image with a Gaussian kernel, $L(x,y,\sigma)= G(x,y,\sigma)*I(x,y)$. Where G(x,y,σ) is the Gaussian function , I(x,y), is the image and L(x,y,σ) is the function which defines the Scale-Space with the variation of σ. Fig.1 shows the first octave of the Scale-Space.

Every octave is created once the previous σ value has been doubled. As discussed by Lowe in [9] every octave was divided into a number of intervals defined by k=21/s with s = 50 and with an initial σ of 0.6. The detection of extrema is performed by doing a search in a 3x3 neighborhood centered in every point in the current scale as well as the same neighborhood in the adjacent scales in the DoG set of data.

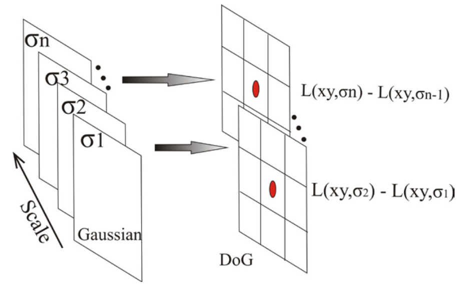

Figure 1. Generation of Scale-Space (left) and the Difference of Gaussian (DoG) used to detect the extrema key points (right).

Key point localization: An interpolation of the maxima is performed by fitting a 3D quadratic function as presented by Bay in [10].

For every image being analyzed, a set of points in x and y coordinates is extracted and then interpolated. This data is used to track the movement of pixels through the movie.

D. SURF

This detector is based on the Hessian matrix. Salient features of the image are detected as locations in which the determinant of the Hessian matrix is maximum, the implementation is based in the description done by Bay in [10].

The Scale-Space is created by applying Gaussian based filters at different scales, after which the Hessian Matrix is calculated for every scale (σ). To speed the algorithm the Gaussian second order derivatives at the different scales were discretized as proposed by Bay in [10]. This representation allows the utilization of integral images for the rapid calculation of the convolution operation using box filters, the size of the box filters increasing as σ increases. The Hessian matrix elements correspond to the convolution of the second order derivative in the different directions and the image. These convolutions are approximated by box filters.

For this detector each octave contains the result of 4 convolutions at 4 different scales and a total of 3 octaves are defined. The extrema detection is performed by detecting the maximum of the determinant of the Hessian matrix. An approximation for the determinant is calculated in [10] as $det(H) \approx D_{xx}D_{yy}-(kD_{xy})^2$.

The value k is a weighted value that was set to 0.9. Once the determinant is calculated for every scale, the extrema detection is done inside a 27-pixel neighborhood of a sample point. The candidates detected are then interpolated with the method shown in [10].

The orientation for each candidate is calculated using Haar wavelets centered in each candidate point. This is calculated locally and did not have a significant influence on the result. A set of key points and an orientation vector is calculated for every image. Finally, the matching process is based on the comparison between the descriptors calculated for the base image and the image being registered.

E. Local descriptors and matching

A local descriptor has to be calculated for every key point calculated in the images. The SIFT descriptor corresponds to a 128 element vector that contains the gradient information of an 8x8 region centered at the key point, this region is divided into four square subregions and one histogram is calculated for every region. The histograms are then concatenated to finally calculate the resulting vector.

The SURF descriptor is calculated for the same subregions described above, Haar wavelet responses in the x and y directions are calculated for every subregion. The final descriptor is a 64 element vector and contains information of the sum of the Haar wavelets in both directions.

III. RESULTS

SIFT and SURF descriptors were calculated for four different data sets and salient points were found successfully with both operators, although a larger number of points were found using the SIFT operator. No additional interventions or starting points were necessary. Fig. 2 shows the performance of both algorithms in the keypoints detection, both algorithms. SIFT produces a more uniformly distributed set of points which predicts a better behavior of this algorithm for the calculation of the transformation required to register the images.

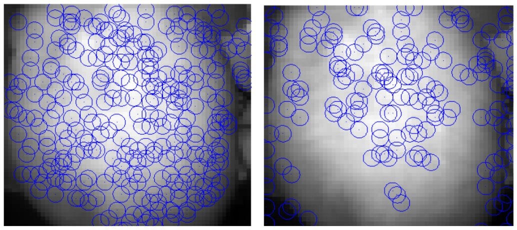

Figure 2. Key points detected with SIFT (left) and SURF (right), both images present keypoints calculated for the same frame in the data set. The actual keypoints are in the center of each circle.

In order to improve the behavior and the speed of the SIFT operator the contour of the heart was detected in every image. Only points located inside the contour were considered candidate points and the descriptor was calculated only for those points.

The matching process was done with the descriptors from both operators; selected results are shown in Fig. 3. During this process the descriptor of one point in the base image is compared to every descriptor in the image to be registered; the euclidean distance between the descriptor vectors is calculated and the point corresponding to smallest distance is chosen as a match. Since one descriptor can have several matches in the second image, a ratio is determined between the distances for the two nearest neighbors. This ratio is compared to a threshold and if it is below threshold the match is accepted. This ratio was 0.9 for the SURF application and 1E-6 for SIFT. The difference in the definition of the descriptors accounts for the difference in the numerical value of the thresholds between SIFT and SURF.

This strategy jointly with previous information about the motion and the nature of the images provide a helpful tool to eliminate mismatches. The movement of a pixel inside the image has to be small, since no contraction can result in a large displacement. With this information a restriction was placed to maintain the motion in a 6x6 pixel ratio. An example of a mismatch can be observed in Fig. 3 (bottom) a black square was placed in the position of the mismatch. This kind of spurious movement can affect the calculation of the final transformation and were controlled with the restriction described.

For the images shown in Fig. 3 a total of 283 points were calculated with SIFT for the base image and 251 a points for the second image, it was possible to track the movement of 107 pixels in the image and after that 11 points were discarded by the physical movement restriction. For SURF a total of 180 points were calculated for the base image and 197 were calculated for the image to be registered, from those points a total of 87 points from the base image were tracked with a valid threshold ratio and 39 points were rejected once the movement restriction was applied.

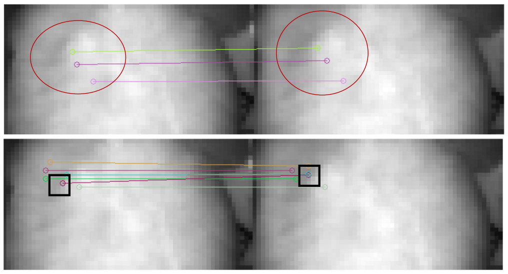

Figure 3. Selected matched point for SIFT (up) and SURF (down) between the base frame (left) and a distorted frame (right).

In order to test the reliability of the algorithms the base image was translated 5 pixels in each direction and a tracking of the keypoints was performed. SIFT tracked a total of 61 points with an average MSE of 0.059 in pixel units, SURF tracked 137 points with an average MSE of 12.7 in pixel units. The large difference between MSEs shown in the results is due to the presence of outliers and mismatches in the keypoints tracking with SURF.

Fig. 3 presents a good example (red circles) describing the variation of shape, intensity and position of the features being tracked. A number of the keypoints detected in the base image are not tracked because the feature disappears in the frame to be registered.

By observing the movement of the points in different regions inside the images it was possible to describe the motion for different regions throughout the sequence of frames, which gives information to calculate a suitable transformation to register the images.

IV. DISCUSSION

Even when both algorithms seem to present an acceptable tracking of the keypoints, SIFT results in a more even distribution of the points over the surface of the heart during the extrema detection step. This distribution would result in better responses of the transformations for image registration.

The type of transformations to be used in the registration process should consider the substantial variations in the movement for different regions in the heart in a single frame. The keypoints detected with both algorithms showed that a global transformation is not convenient to model the movement of the heart surface. For this application affine, similarity and polynomial transformations were applied globally; due to the type of movement present in the image these transformations did not produce an acceptable result for the registration step. The use of local rigid or non-rigid transform models could better represent the local distortions present in the images [11].

The use of box filtering in SURF adds a desirable speed to the algorithm and in previous works has shown a good approximation to the Hessian matrix. However, due to the lack of contrast of the Optical Mapping images, this approximation can result in an increased percentage of mismatches as shown by the results.

The proposed method showed good performance tracking features that have a significant variation inside the movie. An example is the dark region inside the red circles in Fig. 3; this region was tracked over the set of frames even when variations in shape and area were present.

V. CONCLUSION

Scale-space descriptors were adapted to the application presented, yielding a description of movement of pixels distributed on the heart wall, not only in a specific region of the heart.

The change of shape and disappearance of features throughout the sequence of frames add complexity to the registration problem since the solution cannot be reduced to a transformation defining the translation, rotation or scaling of the same scene. This challenge also shows the potential of the scale-space descriptors to correctly determine the movement of a certain keypoint even when major changes occur.

The detection and analysis of the tracked keypoints showed that one global transformation is unable to describe the motion for different regions in the heart and cannot be applied to register the complete image. However, having a list of the tracked keypoints provides the opportunity to effectively determine a suitable transformation for the different regions of the image and perform a proper registration among the frames.

The presented process allows the description of motion of the different regions of the heart throughout the movie, not restricted to a specific, limited ROI. The tracked keypoints will be the starting points for calculating a transformation model for the registration process. Non-rigid models are a promising approach to such transformation models.

REFERENCES

[1] I. R. Efimov, V. P. Nikolski and G. Salama. "Optical Imaging of the Heart". *Circ Res.* Vol. 95, pp. 21-33, 2004.

[2] A. D. Bachtel, R. A. Gray, J. M. Stohlman, E. B. Bourgeois, A. E. Pollard and J. Rogers. "A Novel Approach to dual Excitation Ratiometric Optical Mapping of Cardiac Action Potentials with Di-4-ANEPPS Using Pulsed LED Excitation". *IEEE Trans. Biomed. Eng.*Vol. 58, No. 7, July 2011.

[3] M. Svrcek,S. Rutherford, A. Chen, I. Provaznik,B. Smaill. "Using Image registration to reconstruct spatiotemporal electrical activity in cardiac optical mapping studies". *Comput Cardiol.* Vol. 36, pp.521-524,2009

[4] G. Rohde,B. Dawant and S. Fong. "Correction of Motion Artifact in Cardiac Optical Mapping Using Image Registration", *IEEE Trans. Biomed. Eng.*Vol. 52, No. 2, Feb. 2005.

[5] M. Svrcek,S. Rutherford, A. Chen, I. Provaznik,B. Smaill. "Characteristics of Motion artifacts in Cardiac Optical Mapping Studies". In *Proc. IEEE EMBS.* 2009, pp.3240-3243.

[6] P. Westerfaard, K. Umapathy, S. Masse, E. Sevapstisidis, J. Asta, T. Farid, K. Nair, S. Krishnan and K. Nanthakumar. "Non-Linear image registration for correction of motion artifacts during optical Imaging of Human Hearts", in proc. CCECE/CCGEI. 2008 , pp. 1729-1732.

[7] A. Nygren, A. Lomax and W. Giles, "Voltage sensitive dye mapping in Langendorf-perfused rat hearts", *AM J Physiol-Heart C,* Vol. 298, pp.892-902, 2003.

[8] B. Romeny. "Introduction to Scale-Space theory: Multiscale Geometric Image Analysis," Utrecht University, The Netherlands, Tech. Rep. ICU-96-21,1996.

[9] D. Lowe. "Distictive image features from scale invariant keypoints,". Int J Comput Vision, Vol. 60 Issue 2, Nov. 2004

[10] H. Bay, A. Ess, T. Tuytelaars and L. Van. "Speed-Up Robust Features (SURF)", *Computer vision and image understanding,* Vol.110, Issue 3,pp. 346-359 Jun. 2008

[11] Zitova B., Flusser J.,"Image registration methods: a survey", *Image and Vision Computing*, Vol. 21, 2003, pp. 997-1000