2016 23rd International Conference on Pattern Recognition (ICPR)
Cancún Center, Cancún, México, December 4-8, 2016

Statistical Consensus Matching Framework for Image Registration

Micha Feigin
Department of Mechanical Engineering
Massachusetts Institute of Technology
Cambridge, MA, USA
Email: michaf@mit.edu

Bryan J. Ranger
Harvard-MIT Health Sciences and Technology
Massachusetts Institute of Technology
Cambridge, MA, USA
Email: branger@mit.edu

Brian W. Anthony
Department of Mechanical Engineering
Massachusetts Institute of Technology
Cambridge, MA, USA
Email: banthony@mit.edu

***Abstract*—A common method for image alignment in computer vision is finding the maximum consensus transformation for a set of features in the images. This is commonly done using randomized methods such as RANSAC.**

While relatively robust when strong features are involved, these methods do not deal well with ambiguous features where maximum likelihood does not provide the best match between the images, a common case with modalities such as medical ultrasound, thermal imaging and cross modality registration. They also do not inherently allow for the application of external knowledge regarding possible configurations to aid in the registration.

In this paper we present a novel statistical framework for maximum consensus image alignment which is both robust in the presence of weak features (features not providing one-to-one matches) while at the same time providing an inherent natural ability for integrating external knowledge. Our methods is able to collect information not only from finding good matches, but also from improbable and partially ambiguous matches.

We demonstrate our framework in the context of medical ultrasound image registration. In our test cases, our method succeeded where other state of the art methods we compared to failed to provide satisfactory results with over 17% of the samples.

I. Introduction

One of the most popular methods used in computer vision for the purpose of image registration is maximum consensus. Given a pair of images to align and a set of features in each image, we look for a transformation θ that best matches the largest number of features with a residual up to some threshold ϵ.

We can roughly split the problem on the one hand into rigid or locally rigid transformation (translation, rotation, affine) and deformable registration (optical flow [1], dense SIFT flow [2]). On the other hand we have sparse versus dense registration [3], [1]. In both cases, various feature spaces can be used, some common ones including sum of square differences (SSD), correlation (both of which are based on image space patches), scale-invariant feature transform (SIFT) [4] and Speeded Up Robust Features (SURF) [5].

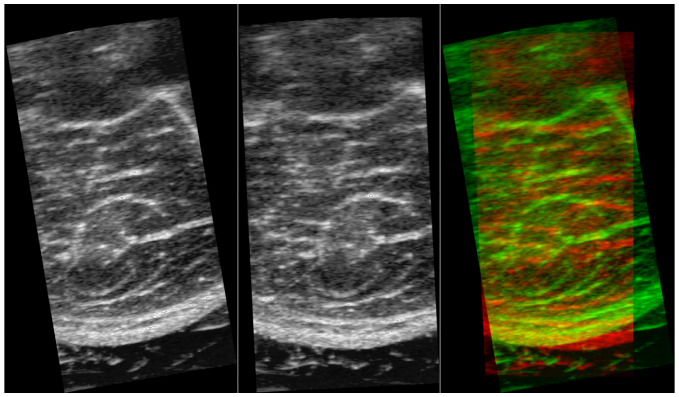
For the case of rigid transformations, due to performance considerations, generally a randomized approach is used. Of these, random sample consensus (RANSAC) [6] has been the dominant approach, along with various proposed improvements [7], [8], [9]. The underlying idea is to randomly choose a minimal set that defines the assumed transformation. Next, we find the maximal set that agrees with the computed transformation. This process is repeated multiple times. Given enough such iterations, a close to optimal transformation is recovered with a high probability.

The major drawback with randomized approaches is that there is no certainty that the near optimal solution will be found, nor that a given solution is indeed close to optimal. Several approaches have been proposed to find an optimal solution, but most of these are very computationally intensive and prohibitively slow. Some examples include branch and bound [10], [11] or dealing with optimal subsets for specific cases [12], [13]. Improving the performance of near optimal solution with some global guaranties is an ongoing field of research. Chin et. al. [9] for example proposed a solution that improves on the performance by using tree searches.

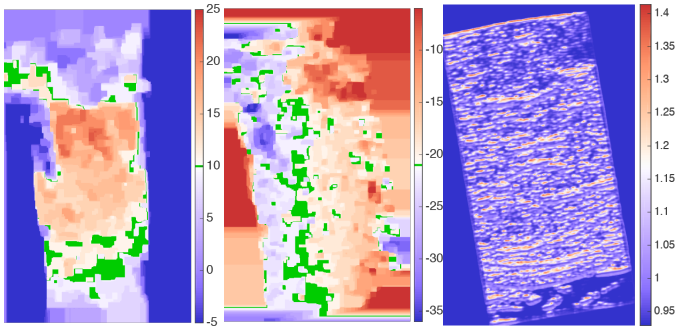
Even ignoring performance, all of these methods suffer from several major inherent drawbacks hampering both robustness as well as flexibility

1. First and foremost, there is no intrinsic way to incorporate external information with regards to the possible configurations. An example being the ability to utilize information from (possibly low accuracy) camera motion tracking.
2. There is an underlying assumption that the maximum likelihood match is in fact the correct one. While this is generally true with high quality images containing strong features, for images resulting from sources such as medical ultrasound imaging, thermal imaging and cross modal registration this is generally not the case. These images tend to lack high quality features, and those that do exist often pose high ambiguity in the matches. Fig 1 shows an example of the issue for medical ultrasound imaging. In this case we see a large number of local and near global optimal points in the feature matching and it is not clear which should be chosen as the correct match.
3. Ambiguous feature matches such as aligning two infinite lines are difficult or impossible to utilize for information.

978-1-5090-4846-5/16/$31.00 ©2016 IEEE



(a) Reference (b) Template (c) L-K Failed alignment



(d) Δx for SIFT flow (e) Δy for SIFT flow (f) Sample SIFT distance image

Fig. 1: The problem with computing alignment with ultrasound images. (a) and (b) show the reference and template images. (c) shows the failed alignment resulting from running the Lucas-Kanade method with image (a) in red and (b) in green. (d) and (e) show the x and y offsets computed using the dense SIFT flow with the green pixels denoting features with correct alignment. (f) shows a typical distance image comparing a template feature to the reference image (lower values denote a better fit).

4. There is no usage of information that can be gained from mismatches, i.e using features that match badly for a given transformation (low probability transformations).

In this paper we propose a new, feature space and metric agnostic, statistical matching approach for rigid alignment where in this first work we demonstrate the results in the context of recovering translation. A naive extension towards handling rotation as well is straight forward, if numerically expensive. We leave research into efficient recovery of more complex transforms for future work.

This framework is targeted at the drawbacks previously mentioned. We construct an alignment confidence map by computing a probability distribution per offset. The confidence for each configuration can both increase based on good matches as well as decrease based on bad matches.

Using such as statistical framework also allows us to combine any form of external information that can be expressed as an expectation distribution with regards to possible outcomes.

We demonstrate the results of our method in the context of medical ultrasound imaging, and specifically, reconstructing a 360 degree tomographic scan of the human leg in the context of prosthetic design [14], [15]. For this use case, all other methods we applied have failed to provide satisfactory results with over 17% of the sample imaging pairs resulting with large registration errors.

II. Method

For the purpose of this work, we use the definitions set forth by Albert Tarantola [16] for disjunction and conjunction of probabilities. Given a set of n probability distributions $p_1, \ldots, p_n$ let us define two new probability distributions: the disjunction and the conjunction probability distributions respectively

$$\begin{aligned} (p_1 \vee \cdots \vee p_n)(x) &= \frac{1}{n}\left(p_1(x) + \cdots + p_n(x)\right) \\ \frac{(p_1 \wedge \cdots \wedge p_n)(x)}{\mu(x)} &= \frac{1}{\nu}\frac{p_1(x)}{\mu(x)} \cdots \frac{p_n(x)}{\mu(x)} \end{aligned} \tag{1}$$

where ν is the normalizing constant

$$\nu = \int_\Omega \frac{p_1(x)}{\mu(x)} \cdots \frac{p_n(x)}{\mu(x)} \, \mathrm{d}x \tag{2}$$

and $\mu(x)$ is the homogenous probability distribution. The homogenous probability distribution is defined when the manifold has a natural notion of *volume* (see [16] for more details). For the case of images in standard cartesian space, this is set to be a constant, but may be set to non-uniform distributions for data that is mapped to more complex manifolds such as polar or spherical coordinates.

These two probability distributions map roughly to *or* and *and* in multivariate logic.

We found the use of conjunction of probabilities to be more robust than disjunction. One way to look at this is that there is a preference to mutual agreement (*and*) rather then averaging where the strong contender (or outlier) can tip the scale. As we will see later (Sec II-C) through a numerical modification, the two are in some sense equivalent, although conjunction has a strong outlier suppressing effect.

A. Probability distribution for registration

For the purpose of registration, we treat the first image as reference and the second image as template. For each feature point y in the template image we can compute the distance function with regards to every point x in the reference image as

$$d_y(x) = \|f_r(x) - f_t(y)\| \tag{3}$$

Here f_r and f_t denote the feature vector in the reference and template images respectively, under an appropriate norm. In the case of SSD features f would denote image patches using the squared l_2 norm. For SIFT feature vectors we use inverse cosine of the inner product (angle).

We take each such match as a the reciprocal of an unnormalized probability distribution. Thus, the given probability distribution is

$$\begin{aligned} p_y(x) &= \frac{1}{\nu_y} \frac{1}{\|f_r(x) - f_t(y)\|} \\ \nu_y &= \int_\Omega \frac{1}{\|f_r(x) - f_t(y)\|} \, \mathrm{d}x \end{aligned} \tag{4}$$

where again ν_y is a normalizing constant.

Note that we take this probability distribution in the general sense as there can be a division by zero when features match exactly, such is the case when matching an image to itself. One solution is to add a small constant offset that can be interpreted as allowing for some uncertainty even in the case of an exact match.

Rather than taking the maximum likelihood solution per feature, which is the common approach, we opt to use the entire probability distribution. The idea is twofold

1. Especially in the case of images suffering from low quality features and / or features that change with angle of view, such as bones in ultra-sound imaging, the maximum likelihood solution is often ill posed, suffering from multiple local optimal points. It is not clear in this case that the global optimum is indeed the best match.
2. There is an opportunity to gain information from bad matches and not only from the good ones. That is, with each feature, we wish not only to gain information from good matches, but also to suppress offsets matching bad matches.

In practice, we do not use all the probability distributions, but rather randomly select a small set of distributions to work with. Specifically, the examples were made with 300 randomly selected feature points.

B. External constraints

By external constraints we mean any external information that does not come directly from the images. Such information can include a rough estimation as to the location of the camera. In our case for example, we know the position of the ultrasound probe, but as the subject is not restrained, there is some relative unknown motion of the subject with respect to the probe.

Such external information can be applied by defining an appropriate probability distribution. Although the two are essentially the same under this framework, we can distinguish between "hard" and "soft" constraints.

Hard constraints mean that we only know the set of valid configurations, but do not know that some options are more likely than others. This can be interpreted as a conditional probability. That is, we want to compute the conditional probability $P\left(x|B\right)$ where B is the domain of validity. This can be defined via conjunction of probabilities as follows (see [16])

$$P\left(A|B\right) = \left(P \wedge M_B\right)(A) \tag{5}$$

In this case M_B is the restriction of the homogenous probability to the event B. The resulting distribution is

$$\mu_B(x) = \begin{cases} k\mu(x) & \text{if } x \in B \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

with k a normalizing constant.

Note that for the common case where μ is a constant, this reduces to the case of multiplying the probability distribution by the support function χ_B, and normalizing.

Soft constraints mean that we know the probability distribution with regards to the expected position. A simple case would be a normal distribution around an expected target point. This case can be simply described again as conjunction of probabilities.

C. Log distribution

The conjunction of a large number of probabilities is numerically unstable, as we are dealing with a large number of multiplication. One solution is to work with log distribution instead

$$\begin{aligned} L\left(x\right) &= \log\left(\frac{1}{\nu}\frac{\left(p_1 \wedge \cdots \wedge p_n\right)(x)}{\mu(x)}\right) \\ &= \log\left(\frac{1}{\nu}\frac{p_1(x)}{\mu(x)} \cdots \frac{p_n(x)}{\mu(x)}\right) \\ &= \log\left(p_1\right) + \cdots + \log\left(p_n\right) - \log\left(\nu\right) - n\log\left(\mu\right) \end{aligned} \tag{7}$$

Assuming that the homogenous probability distribution is uniform then $\log\left(\nu\right) + n\log\left(\mu\right)$ is a constant. As we are looking for the maximum likelihood of the conjunction we can neglect that part, resulting with

$$L\left(x\right) = \sum_{i=1}^{n} \log\left(p_i\right) \tag{8}$$

Ignoring for the moment that the individual log distributions are not probability distributions (due to scaling), we see that the log distribution of the conjunction is equivalent to the disjunction of the log distributions (up to an additive constant due to scaling). This gives an intuition as to why the conjunction is more stable, as the logarithm compresses large positive spikes, reducing the effects of strong false positives (outlier suppression).

D. Aligning probabilities

A final application note is the issue of alignment of probability distributions. Due to the finite size of both template and reference images, two features from the template image cannot convey the same range of transformation parameters. This is due to the fact that some transformations map the template features outside of the reference image. This scenario can be seen in Fig 2, where we see that the probability distributions only partially overlap. As a result, the individual probability distributions do not cover the entire domain and need to be padded. We use zero padding for both disjunction and conjunction of the log distribution.

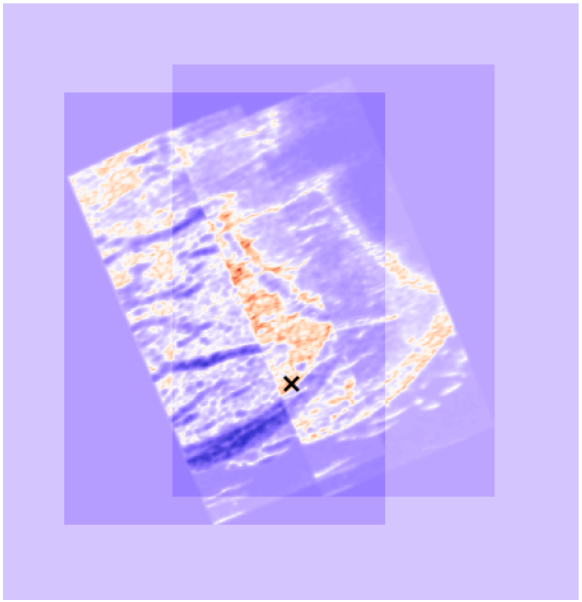
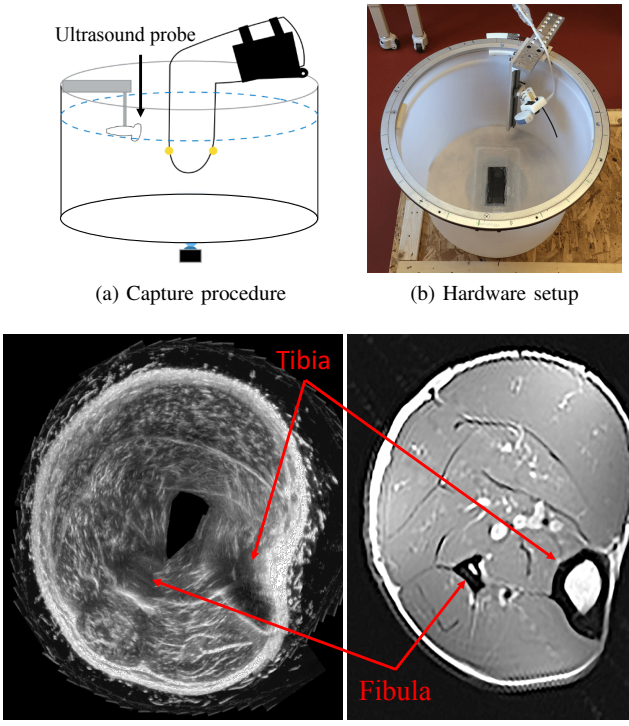


Fig. 2: Here we see how probability distributions for displacement based on different features only partially overlap as due to the finite size of the image, some offsets map to coordinates which are outside the image.



(a) Capture procedure (b) Hardware setup

(c) Comparison of US (left) and MRI (right) results

Fig. 3: This figure presents the capture setup (a and b) as well as a comparison of the US tomographic reconstruction (c, left) to the MRI image (c, right) of approximately the same slice. Images are of the lower leg. US image has been reconstructed from 72 individual images.

For disjunction, this creates a small bias towards zero offset, although in our case we did not find that it warrants special handling.

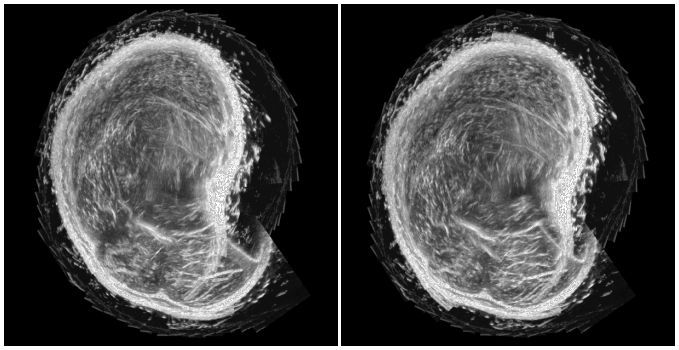
For conjunction, this complicates things due to two issues. The first is that a value of zero in the log probability maps to one in the original distribution, which leads to the second problem. The log distributions may be negative. If not careful, zero padding may create a bias towards large offsets. The solution we used is to scale all distributions uniformly to reach a mean value on the order of one (zero in the log distribution). Scaling of the distribution can also be interpreted as scaling of the domain as a probability distribution needs to integrate to one. One needs to make sure though to use the same value for all distributions. In our case we found that dividing the original distribution by the mean (regularization) and then multiplying by the number of pixels in the domain (scaling) provided a good results.

III. Results

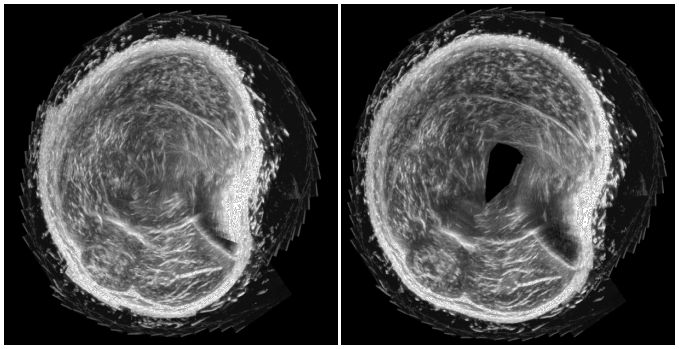
Our sample problem depicts a 360 degree tomographic ultrasound scan of the lower leg. Capture was taken of a healthy human subject (under approval from the MIT COUHES office). Figures 1a and 1b show two sample images from the scan (reference and template images). The scanning setup is depicted in Fig 3a with an image of the actual setup shown in Fig 3b. An ultrasound probe is mounted onto a rotational stage and is rotated around the subject's lower leg to collect 72 images at 5 degree increments. The angle and location of the probe is known, but the subject is unrestrained (both for comfort as well as to avoid physical distortion of the extremity during measurements). This results with relative motion of the subject requiring compensation to achieve proper reconstruction. Our problem is reduced to rigid registration between pairs of ultrasound images (see [14], [15] for more information on the setup and medical problem statement). Fig 3c depicts a full correct reconstruction based on our method from the ultrasound images (left) and the corresponding MRI image (right). The dark area in the center of the ultrasound image is due to lacking penetration in the capture, meaning that the scan does not cover the whole volume of the leg.

In figure 4 we can see the results of attempted full reconstruction using several methods. As a reference implementation we used the image alignment toolbox [17]. Fig 4a is the attempted reconstruction using the Lucas-Kanade method [18], [19]. Fig 4b shows the same using the enhanced correlation coefficient (ECC) [20]. Both produce large errors on 13 and 12 of the image pairs respectively (approximately 17%). SURF based registration [5] as implemented by both MATLAB as well as the image alignment toolbox failed to find any useful feature on most of the images. Figures 1d and 1e depicting the dense SIFT flow [2] show that it is also impossible to figure out the correct offset from this method (green depicts the correct offsets), at least not without extra filtering heuristics and noting that correct offsets do not correlate with good features.

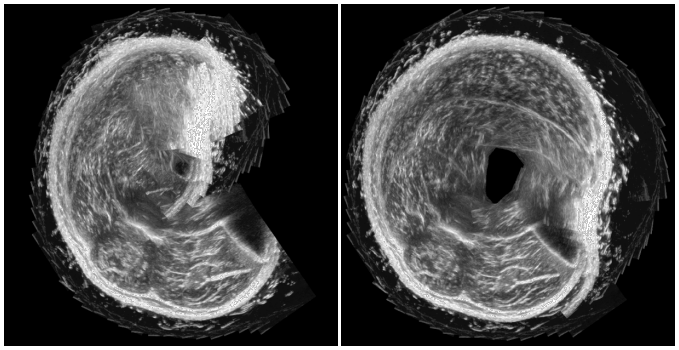
To present our method, we implemented recovery based on 300 random features per image pair. We performed recovery



(a) Lucas-Kanade (b) ECC



(c) SIFT, disjunction (d) SIFT, conjunction



(e) SSD, disjunction (f) SSD, conjunction

Fig. 4: Full reconstruction comparing Lucas-Kanade registration (a), ECC registration (b) along with several variations of our method. (c) and (d) show the alignment based on SIFT features baed on disjunction and conjunction of probabilities respectively. (e) and (f) show the same based on SSD features. The only correct reconstruction is (d), with (f) almost as good with only one significant registration error. All other options have multiple errors, also presented in Fig 5.

using both disjunction (logical or) and conjunction (logical and) of probabilities. No heuristics were used in choosing the features other than ensuring that the chosen features are within the domain of interest (i.e image rather than mask features resulting from image rotation). As can be seen in Fig 4d, our method of statistical based offset recovery using SIFT features based on the conjunction of probabilities achieves perfect (or near perfect) reconstruction. Doing the same with sum of square difference features (SSD) using 8×8 sized patches, as depicted in Fig 4f achieved almost the same results, producing an erroneous reconstruction for only two image pairs. Using Disjunction of probabilities (Figures 4c and 4e) proved to produce erroneous results and is presented for completeness.

In Fig 5a we use our recovery based on the statistical method using SIFT features and conjugation of probabilities as ground truth. This reconstruction has been deemed by visual inspection to be as good as can be achieved. The distance (error) per recovered translation for each image pair as compared to each of the other recovery methods depicted in Fig 4 is plotted as measured in pixels. As can be seen, registration as recovered by using SSD features and our statistical method and conjugation of probabilities is the only one that comes close, with two large errors for frames 29 and 71. All of the rest of the methods suffer from multiple failed frames, where at least 17% of the frames contain large registration errors. Fig 5b depicts the matching recovered position for each one of the sets appearing in Fig 4, with the solid black line showing our reference ground truth.

Finally, in Fig 6 we can see example probability distributions for SIFT based registration for disjunction of probabilities (Fig 6a) and conjunction of probabilities (Fig 6b). Fig 6c shows the resulting registration. The images in this case are the same as depicted in Fig 1.

IV. Conclusion and future work

In this work we presented a novel statistical framework for recovering maximal consensus matching for the purpose of rigid image registration. This framework is not only much more robust than random consensus matching as it is able to extract information from highly ambiguous matches as well as mismatches, it also allows for the inclusion of external information such as camera and object tracking in an intrinsic way to the framework.

As future work, we look to expand the method efficiently to more general transformations.

References

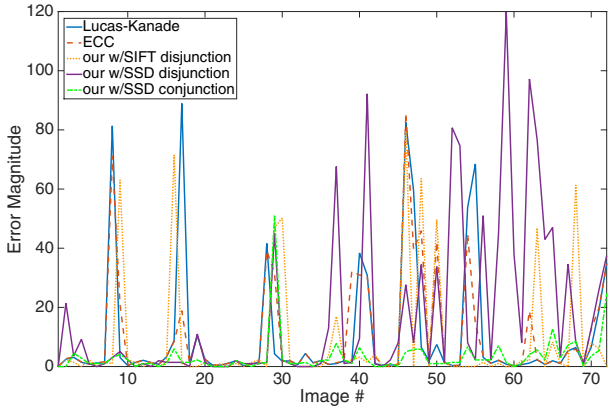
[1] N. Paragios, Y. Chen, and O. Faugeras, *Handbook of Mathematical Models in Computer Vision*. Springer, 2006.

[2] C. Liu, J. Yuen, and A. Torralba, "Sift flow: Dense correspondence across scenes and its applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 5, pp. 978–994, 2011.

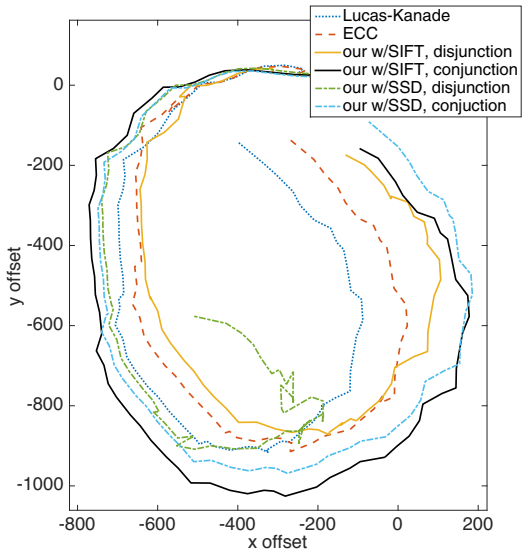
[3] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge university press, 2003.

[4] D. G. Lowe, "Object recognition from local scale-invariant features," in *ICCV*, vol. 2, no. 8, 1999, pp. 1150–1157.

[5] H. Bay, A. Ess, T. Tuytelaars, L. Van Gool, L. V. Gool, H. Baya, A. Essa, T. Tuytelaarsb, and L. Van Goola, "Speeded-up robust features (surf)," *Computer vision and image understanding*, vol. 110, no. 3, pp. 346–359, 2008.

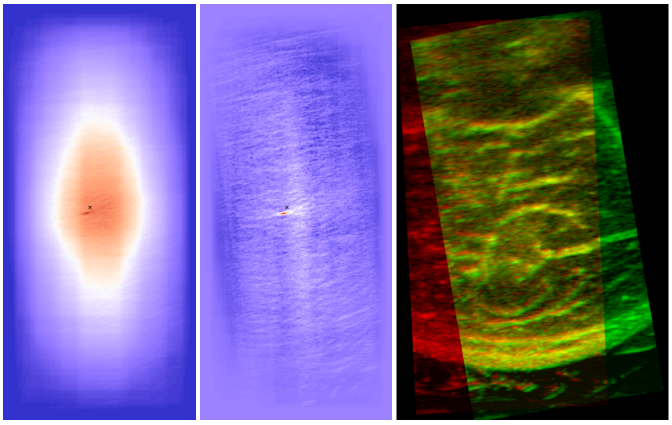


(a) Registration error



(b) Recovered positions

Fig. 5: In (a) we took the recovery based on our statistical method using SIFT features and conjunction of probabilities as the ground truth, as it provides the best alignment of the entire set, and plot the registration error per image pair measured in pixels for each one of the other methods. Conjunction of probabilities with SSD features provides the best comparative result with only two relatively large errors on frames 29 and 71. All other registration variations results with multiple bad registration frames, on the order of 17% bad frames. Fig (b) shows the recovered positions for each of the images, with the ground truth case in solid black.



(a) Disjunction (b) Conjunction (c) Registration

Fig. 6: Sample statistical distributions used for computing alignment. (a) Shows the statistical distribution using disjunction based on 300 SIFT features (b) shows the same for conjunction. (c) Shows the resulting registration based on only 150 random SIFT features. These are for the images shown in Fig 1 which result in failed registration with Lucas-Kanade and ECC registration as well as SIFT and SURF based registration.

[6] M. a. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applicatlons to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381 – 395, 1981.

[7] P. H. S. Torr and C. Davidson, "IMPSAC: Synthesis of importance sampling and random sample consensus," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, pp. 354–364, 2003.

[8] P. H. S. Torr and A. Zisserman, "MLESAC: A new robust estimator with application to estimating image geometry," *Computer Vision and Image Understanding*, vol. 78, no. 1, pp. 138–156, 2000.

[9] T.-j. Chin, P. Purkait, A. Eriksson, and D. Suter, "Efficient globally optimal consensus maximisation with tree search," in *CVPR*, 2015, pp. 2413–2421.

[10] Y. Zheng, S. Sugimoto, and M. Okutomi, "Deterministically maximizing feasible subsystem for robust model fitting with unit norm constraint," in *CVPR*, 2011, pp. 1825–1832.

[11] H. Li, "Consensus set maximization with guaranteed global optimality for robust geometry estimation," in *ICCV*, 2009, pp. 1074–1080.

[12] O. Enqvist, E. Ask, F. Kahl, and K. Aström, "Robust fitting for multiple view geometry," in *ECCV*, 2012, pp. 738–751.

[13] C. Olsson, O. Enqvist, and F. Kahl, "A polynomial-time bound for matching and registration with outliers," in *CVPR*, 2008.

[14] B. Ranger, M. Feigin, N. Pestrov, X. Zhang, V. Lempitsky, H. Herr, and B. W. Anthony, "Motion compensation in a tomographic ultrasound imaging system: Toward volumetric scans of a limb for prosthetic socket design," in *EMBC*, Milan, Italy, 2015.

[15] B. Ranger, M. Feigin, X. Zhang, A. Mireault, R. Raskar, H. Herr, and B. W. Anthony, "3d optical imagery for motion compensation in a limb ultrasound system," in *SPIE Medical Imaging*, San Diego, California, USA, 2016.

[16] A. Tarantola, *Inverse Problem Theory and Methods for Model Parameter Estimation*. Siam, 2005.

[17] G. Evangelidis, "Iat: A matlab toolbox for image alignment," 2013.

[18] S. Baker, R. Gross, T. Ishikawa, and I. Matthews, "Lucas-kanade 20 years on : A unifying framework : Part 2," *International Journal of Computer Vision*, vol. 56, no. 3, pp. 221–255, 2004.

[19] S. Baker and I. Matthews, "Lucas-kanade 20 years on : A unifying framework : Part 1," *International Journal of Computer Vision*, vol. 56, no. 3, pp. 221–255, 2004.

[20] G. D. Evangelidis and E. Z. Psarakis, "Parametric image alignment using enhanced correlation coefficient maximization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 10, pp. 1858–1865, 2008.