n-SIFT: n-Dimensional Scale Invariant Feature Transform

Warren Cheung and Ghassan Hamarneh

***Abstract*—We propose the n-dimensional scale invariant feature transform (n-SIFT) method for extracting and matching salient features from scalar images of arbitrary dimensionality, and compare this method's performance to other related features. The proposed features extend the concepts used for 2-D scalar images in the computer vision SIFT technique for extracting and matching distinctive scale invariant features. We apply the features to images of arbitrary dimensionality through the use of hyperspherical coordinates for gradients and multidimensional histograms to create the feature vectors. We analyze the performance of a fully automated multimodal medical image matching technique based on these features, and successfully apply the technique to determine accurate feature point correspondence between pairs of 3-D MRI images and dynamic 3D + time CT data.**

***Index Terms*—Biomedical image processing, difference of Gaussian, feature extraction, image matching, medical images, scale invariant feature transform (SIFT).**

I. Introduction

ESTABLISHING correspondence between feature points in a pair of images is important for landmark-based image registration, and for building statistical models of shape and appearance [1]–[3]. Applications include stitching multiple images of the same patient from a session into a unified whole, aligning images taken of the same patient at different times or modalities, or aligning one image time-series with another. The scale-invariant feature transform (SIFT) produces stable features in 2-D images [4], [5]. We generalize SIFT to n-dimensional images (n-SIFT), and evaluate our extensions in the context of medical images. n-SIFT locates positions that are stable in the image, creating a unique feature vector, and matches the feature vectors between two scalar images of arbitrary dimensionality.

At the extrema of the difference of Gaussian (doG) scale space, feature vectors are constructed from local image gradients weighted by the distance from the feature position. Three variations of n-SIFT are proposed and evaluated—a global weighted histogram-based feature vector, the histogram-based feature reoriented in the direction of local gradients, and our extension of the standard 2-D SIFT to n dimensions which we refer to as the n-SIFT feature vector. We use n-dimensional (nD) gradient vectors with $n-1$ directions and a magnitude, summarized using a weighted multidimensional histogram. As well, the 128D SIFT feature vector has been generalized to a 2^{5n-3}D feature vector (Section III-B).

Manuscript received August 16, 2008; revised May 02, 2009. First published June 05, 2009; current version published August 14, 2009. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Peter C. Doerschuk.

W. Cheung is with the Bioinformatics Program, University of British Columbia at the Centre for Molecular Medicine and Therapeutics, Vancouver, BC V5Z 4H4 Canada (e-mail: wcheung@cmmt.ubc.ca).

G. Hamarneh is with the Medical Image Analysis Lab, Simon Fraser University, Burnaby, BC V5A 1S6 Canada (e-mail: hamarneh@cs.sfu.ca).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2009.2024578

II. Related Work

Detection of features in images is a vast research area, encompassing many methods to both detect feature points of interest, as well as describing the features in a manner to facilitate comparison of features between related images. Existing strategies for detecting salient features include edge and corner detection, such as Canny edge detection [6], Harris/Plessey corner detection [7], Shi and Tomasi corner detector [8], SUSAN corner detector [9], and FAST corner detector [10]. The use of doG in the n-SIFT approach evokes similar work in blob detection, such as the MSER blob detector [11] and work by Lindeberg [12], [13]. Most pertinent to this work, however, are the developments on the use of doG and SIFT features.

DoG image filtering has been previously used to successfully identify features in breast thermograms for early detection of breast cancer [14], suggesting that extrema in doG may also indicate salient features. SIFT features and their extensions have been previously applied for object recognition as an easily computed scale-space kernel approximating the scale-normalized Laplacian of Gaussian [4], [5], [15] and have been shown to be most resistant against common image deformations [16].

In the domain of medical image analysis, the robustness of SIFT features has been previously evaluated on 2-D medical images. SIFT features show stability under arbitrary affine transformations [17], [18] when tested on MRI and ultrasound images. A B-spline based interpolation framework using SIFT feature points showed promising performance registering elastically deformed MRI images, although ultrasound images in this case proved more challenging [18]. Improved matching of images across multiple modalities under rotation has been achieved by normalizing gradient magnitudes and restricting gradient orientations to $[0, \pi)$ [19]. SIFT features have been used as part of automated image classification systems using Euclidean distance between SIFT features [20]. The SIFT feature localization has also been used to deterministically localize control points for elastic image registration [21].

A modified version of the SIFT feature has been previously adapted to 3-D, using feature positions detected by Foerstner corners without rotational invariance, and applied to thoracic CT images [22]. The authors verify the technique against synthetic nonlinear deformations as well as clinical data. The modified SIFT feature uses multidimensional histograms to summarize

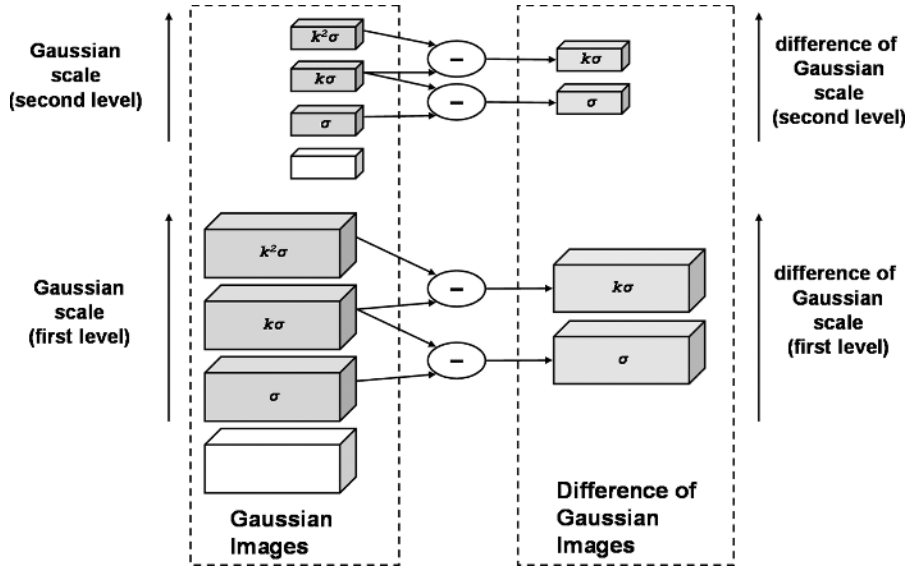

Fig. 1. Image pyramid—Each level is a downsampled image, convolved with Gaussian scales $\sigma, k\sigma, k^2\sigma$ and so on, and doG images from adjacent scales.

gradient directions, but the feature summarizes the gradients at the pixel locations in a $2\times2\times2$ grid. The modified SIFT feature is an intermediate between the histogram feature vector and the n-SIFT feature vector we propose. The 2-D SIFT feature is the $n = 2$ case of the n-SIFT feature we describe in Section III.

In the following sections, we present a detailed description of the three proposed features (Section III), then evaluate the features using 3-D and 4-D medical image data (Section IV) and conclude by discussing the results (Section V). We compare a simple histogram-based feature against the addition of a method for rotational invariance and the n-SIFT method using several test images, and show the effectiveness of n-SIFT at successfully matching points under minor rotational change. A preliminary version of the work was published [23] and an open source implementation has been made available to the research community [24].

III. Methods

The computation of any of the three features consists of two distinct steps: feature localization followed by feature generation. Feature localization involves finding distinctive locations in the image. During feature generation, a feature vector is generated for each of the distinctive locations.

Once features for a pair of images have been computed, correspondences between the feature points in one image with points in the other image can be evaluated. This involves matching the feature vectors from one image to the feature vectors in the other while eliminating ambiguous matches.

A. Feature Localization

The first step in matching points across medical images depends on localizing distinctive points in the image that are relatively robust with respect to the deformations expected and differences between subjects. For all three features, we identify maxima and minima in difference of Gaussian scale space (Algorithm 1).

Algorithm 1 Feature Localization in Difference of Gaussian Scale Space. Takes an image I and generates a set of features F by computing difference of Gaussian images (see Fig. 1). Parameters are $MaxImageLevel$ (number of image levels to compute), σ (sigma for Gaussian Filter) and s (number of Gaussian scales to compute at each level) are specified by the user.

1: $I_0 = I$
2: $F := \emptyset$
3: **for** $i = 0$ to $MaxImageLevel$ **do**
4: $\quad I_{i,0} := I_i$
5: $\quad$ **for** $j = 1$ to $s + 3$ **do**
6: $\quad\quad I_{i,j} := GaussianFilter(I_i, k^j\sigma)$ {Gaussian Filter Image I_i with sigma $k^j\sigma$}
7: $\quad\quad dI_{i,j} := I_{i,j} - I_{i,j-1}$
8: $\quad$ **end for**
9: $\quad$ **for** $j = 1$ to $s + 2$ **do**
10: $\quad\quad pI := FindExtrema(dI_{i,j-1}, dI_{i,j}, dI_{i,j+1})$ {See Section III-A1}
11: $\quad\quad$ **for all** $p \in pI$ **do**
12: $\quad\quad\quad F := F \cup GenerateFeatures(p, I_i)$ {See Section III-B}
13: $\quad\quad$ **end for**
14: $\quad$ **end for**
15: $\quad I_{i+1} := Scale(GaussianFilter(I_i, \sigma), 0.5^i)$
16: **end for**
17: **return** F

To achieve invariance to image scaling, a multilevel image pyramid is created, similar to the one employed by 2-D SIFT [4], [5]. The first image of each successive level (represented by white boxes in Fig. 1) is a linearly interpolated, downsampled version of the Gaussian smoothed image at scale σ (where σ is the standard deviation for the n-dimensional Gaussian) in the previous level (Algorithm 1 , line 15). For our experiments,

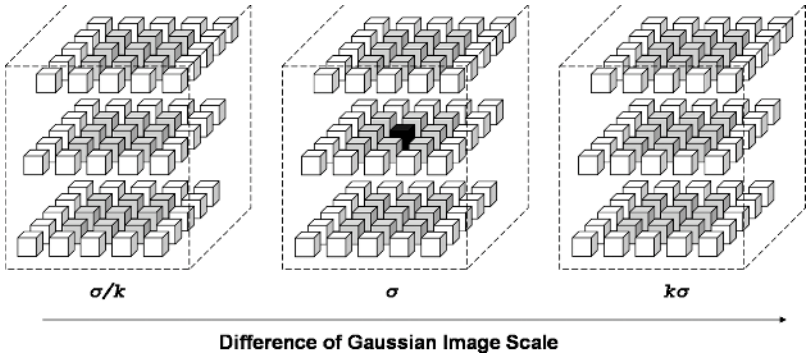

Fig. 2. For 3-D images, an extremum (in black) is the maximum or minimum of the neighbors (shaded) at the same scale, and the corresponding voxels in the scale above and below.

we downsampled the images by half in each dimension at each level, yielding images of the same size as the original image at the first level, a level of images half-sized in each dimension, quarter-sized at the next, and so on.

1) Locate Extrema in the DoG Space: At each level, we examine an approximation of the doG space from σ to 2σ by taking the difference of Gaussian blurred images. To test s doG image scales in a pyramid level, select k such that $k^s = 2$. Starting from the first image at each level, a series of Gaussian blurred images are generated using $\sigma = \sigma,\ k\sigma,\ k^2\sigma$ and so on (lines 5, 6). From each neighboring pair of blurred images at scales $k^j\sigma$ and $k^{j+1}\sigma$, a doG image at scale $k^j\sigma$ is generated by taking the difference of the images (line 7 and Gaussian images of Fig. 1).

Within each pyramid level, we locate extrema (local maxima and minima) in our approximation of doG space (line 10). Each voxel of a doG image (scale $k^j\sigma$) is compared against the neighboring voxels immediately adjacent, orthogonally and diagonally, the corresponding voxel in the scale above $(k^{j+1}\sigma)$ at the same pyramid level, and all the neighbors to that corresponding voxel, and the corresponding voxel in the scale below $(k^{j-1}\sigma)$ and all the neighbors.

Let us refer to a voxel p from the doG image scale $k^j\sigma$ at position $(x_1, x_2, \ldots, x_n)$ as $p_j(x_1, x_2, \ldots, x_n)$.

Definition 1: A voxel $p_j(x_1, x_2, \ldots, x_n)$ is a local extrema iff $\forall_{i_0=-1,0,1}\forall_{i_1=-1,0,1}\forall_{i_2=-1,0,1}\ldots\forall_{i_n=-1,0,1}$,
$|p_j(x_1, x_2, \ldots, x_n)| \geq |p_{j+i_0}(x_1 + i_1, x_2+i_2, \ldots, x_n + i_n)|$
or $\forall_{i_0=-1,0,1}\forall_{i_1=-1,0,1}\forall_{i_2=-1,0,1}\ldots\forall_{i_n=-1,0,1}$,
$|p_j(x_1, x_2, \ldots, x_n)| \leq |p_{j+i_0}(x_1 + i_1, x_2 + i_2, \ldots, x_n + i_n)|$.

For an nD image, there are a total of $3^{n+1}-1$ neighbors to be checked—the 3 width hypercube of voxels excluding the central voxel. Fig. 2 shows an example for 3-D images.

As the scale above and below the current scale are needed to determine whether a voxel is an extremum, $s+2$ doG images and $s+3$ Gaussian images will need to be generated. Finally, only extrema in the doG image having magnitude greater than a threshold T_{doG} are considered feature points.

By computing feature positions and feature descriptors for each level of the image pyramid, we incorporate robustness against image scaling. Note that as we compute multiple image levels for both of the images being matched, the system is equally effective against magnification and reduction, although a significant difference in image size means that the smaller image will have substantially fewer features to match.

B. Feature Generation

At each of the extrema localized, an identifying feature is generated. Regardless of the pyramid level or scale where the extrema was localized, or the feature generation method used, the feature vector generated will be associated with the physical location in the *original* image corresponding to the location of the extrema detected. For our analysis, we investigate three related methods for generating features involving histograms of the local gradient.

The *global weighted histogram feature* (GWH) is a degenerate case of a SIFT feature where only a single histogram is used to summarize the gradients near the located feature point. The histograms summarize the direction of the gradient. In nD, we represent the direction of the gradient via $n-1$ hyperspherical coordinates. For example, in 2-D, one angle—θ, the polar angle, in polar coordinates—is sufficient to describe direction. Similarly, in 3-D, two angles—θ, the azimuthal angle and ϕ, the polar angle, in spherical coordinates—will suffice. In general, we can treat the direction as a point on an n-dimensional unit n-sphere $S^n = \left\{x \in \mathbb{R}^{n+1} : \|x\| = 1\right\}$, described by $n-1$ angular coordinates and where the radial coordinate $r = 1$.

These angles can then be summarized using a $n-1$ dimensional histogram, with b bins spanning 2π radians for each orientation, for a total of b^{n-1} bins. For each voxel in the image, we generate a value as the product of the magnitude of the gradient and a Gaussian centred on the feature position. This value is added to the corresponding bin for the gradient orientation. Although all voxels in the image were considered in these experiments, in practice only voxels within a couple σ of the feature position need to be considered due to the Gaussian weighting and would substantially improve runtime performance.

The histogram for a feature point can then be represented as a b^{n-1} dimensional vector, which is normalized. Normalizing the histogram allows matching of images between different modalities, as noticed previously in 2-D medical images [19], but also makes the feature resistant to image intensity differences between the images.

The *reoriented global weighted histogram feature* (RGWH) is computed in much the same manner; however, it normalizes the direction of the gradients with respect to the highest value bin in the histogram, increasing robustness to changes in orientation. Effectively, the orientation of all the gradient

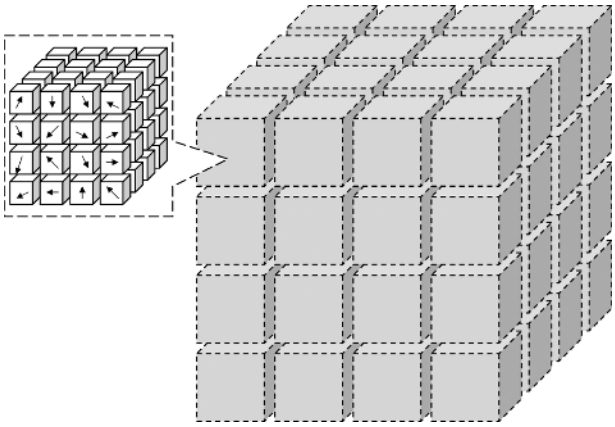

Fig. 3. In 3-D, each of the $4 \times 4 \times 4$ regions (dashed and shaded) summarizes the gradients at $4 \times 4 \times 4$ voxel locations (solid and white).

vectors is rotated such that the highest value bin of the histogram always falls in the same histogram bin. This is implemented by computing the histogram as per the GWH, and then identifying the histogram bin with the highest value—the peak. The index for the peak is first converted to its average orientation direction in hyperspherical coordinates—the peak vector $(\theta'_0, \theta'_1, \ldots, \theta'_{n-1})$. A new histogram is then created by taking each bin in the original histogram, converting its bin index to hyperspherical coordinates $(\theta_0, \theta_1, \ldots, \theta_{n-1})$, rotating it by subtracting the peak vector $(\theta_0 - \theta'_0, \theta_1 - \theta'_1, \ldots, \theta_{n-1} - \theta'_{n-1})$, and then converting this vector to the bin index of the new histogram where we store the contents of the bin. Note that the peak after reorientation will always reside in the bin corresponding to the coordinates $(0, 0, \ldots, 0)$.

The n-SIFT *feature* summarizes a 16^n hypercubic voxel region around the feature position. In contrast with the previous two features which use a single global weighted histogram, the n-SIFT feature divides the local area into 4^n subregions, each using a 8^{n-1} bin histogram to summarize the gradients of the voxels in the subregion, resulting in a $4^n 8^{n-1} = 2^{5n-3}$-dimensional feature vector (Fig. 3). For example, for 2-D $(n = 2)$, the region summarized a 16×16 pixel square , divided into 16 subregions which are each 4×4 pixel squares. In 3-D $(n = 3)$, we have a $16 \times 16 \times 16$ voxel cube, divided into 64 subregions, each summarizing a $4 \times 4 \times 4$ voxel cube. The n-SIFT feature described here implements the key features of the 2-D SIFT feature [5]. The histograms summarize the gradient direction, again weighted by a Gaussian centred at the feature position.

C. Feature Matching

To match histogram(s) generated by any one of the three types of features, we convert the histogram(s) of a feature to a single vector, and compare the l^2 distance between a feature vector in one image against every feature vector of the second image. For feature v, let u be the best match (lowest distance), u' be the second best match, and $d(v, u)$ be the distance between features v and u. We then make sure $d(v, u)/d(v, u')$, is below a threshold T_m and that v is, conversely, the best match for u. This decreases mismatches by removing matches where other features are very close in feature space to the best match.

TABLE I
RUNTIMES WHEN THE TEST TRANSFORMATION IS A SCALE BY 0.8, FOR BOTH THE ORIGINAL IMAGE AND THE TRANSFORMED IMAGE, SHOWING TIME (IN SECONDS) FOR EXTREMA DETECTION ALONE, AND FOR EXTREMA DETECTION, FEATURE GENERATION AND MATCHING FOR EACH OF THE THREE FEATURES ON THE ORIGINAL AND TEST IMAGE

Modality	Size (voxels)	Extrema Detect	GWH	RGWH	n-SIFT
T1	$90 \times 108 \times 90$	101	660	670	410
PD	$90 \times 108 \times 90$	100	730	740	450
T2	$90 \times 108 \times 90$	102	530	530	330

IV. RESULTS AND VALIDATION

We applied these features to 3-D human magnetic resonance imaging (MRI) brain scans and 4-D $(3\mathrm{D} + \mathrm{time})$ canine computed tomography (CT) cardiac scans, showing stability under rotation and scaling transformations. Parameters were set to those used by Lowe [5] and autopano-SIFT [25] as follows: the Gaussian blur for the doG images was performed with $\sigma = 1.5$ and sampling rate of $s = 3$, with T_{doG} for magnitude filtration at 0.0075. For the GWH and RGWH features, we set $b = 36$ and Gaussian $\sigma = 8.0$ for weighting, matching the values previously used for reorienting the SIFT feature vector. For n-SIFT, we set $b = 8$ and $\sigma = 4.0$. The matching threshold T_m was 0.8.

Three 1-mm isotropic BrainWeb simulated brain 3-D MRI data (cropped size $90 \times 108 \times 90$)[1] [26] were used. 4-D CT images of in-vivo canine heart from the Mayo Dynamic Spatial Reconstructor, 0.925 mm isotropic $\times 18$ images per cardiac cycle (cropped size $78 \times 80 \times 88$ voxels $\times 9$ time points)[2] [27], were also used. Extrema were searched through three octaves. The images were transformed using linear interpolation. Runtime on all pairs of sample 3-D images (feature extraction for both images and matching features) was less than 19 CPU minutes for over 874 thousand voxels per image. Runtimes on pairs of sample 4-D images was less than 37 CPU minutes for over 4.9 million voxels per image. Timed tests were run on an Intel Pentium M 1.7-GHz machine with 1 GB of memory. For comparison, runtime of the same pairs of 3-D images on Pentium 1.6 GHz Itanium 2 machines with 64 GB of memory was less than 6 CPU minutes.

A. Evaluating the DoG Extrema Extraction

For the 3-D images, extracting the extrema took 100–102 CPU seconds (CPU s). Generating and matching feature vectors were much more computationally intensive parts. The total runtime with these steps was 530–740 CPU s for the GWH and the RGWH features, with reorientation causing no more than 10 CPU s difference. n-SIFT is the quickest of the features to run, taking a total of 330–415 CPU s, although adding reorientation the n-SIFT feature would negate this by requiring generation of the GWH (Table I).

The extrema for an image and the same image scaled to 0.8 of its original size, rotated by 10°, or both scaled and rotated

[1]http://public.kitware.com/pub/itk/Data/BrainWeb/ (Accessed July 2009).

[2]http://nova.nlm.nih.gov/Mayo/NLMDATA/Dog/ (Accessed March 2006).

TABLE II
NUMBER OF EXTREMA LOCALIZED FOR THE BRAINWEB 3-D IMAGE, AND THE SAME IMAGE DOWNSCALED BY 0.8. THE MATCHES GENERATED BY THE GWH, RGWH, AND n-SIFT FEATURES ARE COMPARED

Modality	Extrema (Orig.)	Extrema (Scaled)	GWH	RGWH	n-SIFT
T1	449	275	107	73	144
PD	516	286	110	83	142
T2	363	163	81	66	90

TABLE III
STABILITY OF EXTREMA FOR TRANSFORMED 3-D MRI IMAGES. MATCHES ARE THE MEAN OF THE FRACTION OF EXTREMA CORRECTLY, UNDER VARIOUS ERROR CRITERIA IN VOXELS. THE TRANSFORMATIONS ARE SCALING BY A FACTOR OF 0.8, ROTATING BY 10°, OR BOTH SCALING AND ROTATING

Transform	Extrema		Match	Match	Match
	(Orig.)	(Trans.)	< 1.5	< 3.0	< 7.5
scale	443	241	0.474	0.680	0.915
rotate	443	427	0.480	0.603	0.773
rotate+scale	442	223	0.399	0.544	0.800

TABLE IV
MEAN FRACTION OF GWH FEATURES MATCHED ACCURATELY (UNDER ERROR IN VOXELS)

Image Type	Transform Type	Matches (< 1.5)	Matches (< 3.0)	Matches (< 7.5)
3D MRI	scale	0.572	0.839	0.968
3D MRI	rotate	0.428	0.462	0.557
3D MRI	rotate+scale	0.172	0.211	0.333
4D CT	rotate	0.625	0.625	0.625

were compared. For each extremum located in the scaled image, the closest (ℓ_2 distance) corresponding extremum in the original image is located, and the margin of error (ℓ_2 distance) computed. Table II shows half to two-thirds of GWH features in the scaled image have no corresponding feature in the original image within 1.5 voxels. n-SIFT features, however, are the most discriminatory, with potential matches for more than half the features. Table III shows that under all error margins and transformations tested, a significant portion (over 0.39) of the extrema potentially matched, and over 0.77 of extrema matching if the error bound is relaxed to 7.5 voxels.

B. Evaluating Feature Matching

We tested the stability and uniqueness of the features generated in the context of medical images under scaling of 0.8 of all axes and rotation of 10° along a single axis. Matches were evaluated by Euclidean distance of the matched point in the original image from the true corresponding point.

1) Global Histogram Feature Evaluation: Table IV shows the (GWH) feature, for 3-D MRI data, is successful under the scale transform alone, correctly matching over 0.95 of the features at the largest margin of error (7.5 voxels). However, when rotation is considered, the accuracy plummets to 0.56, but even at the most stringent error margin, the accuracy is still at a not insignificant 0.43. When considering both scaling and rotation, accuracy drops well below 0.5 for all images at all error margins. However, this feature performs fairly well (> 0.6 accuracy at all error margins) on rotated 4-D CT data.

TABLE V
MEAN FRACTION OF RGWH FEATURES MATCHED ACCURATELY (UNDER ERROR IN VOXELS)

Image Type	Transform Type	Matches (< 1.5)	Matches (< 3.0)	Matches (< 7.5)
3D MRI	scale	0.480	0.675	0.840
3D MRI	rotate	0.541	0.671	0.792
3D MRI	rotate+scale	0.341	0.480	0.641
4D CT	rotate	0.667	0.667	0.778

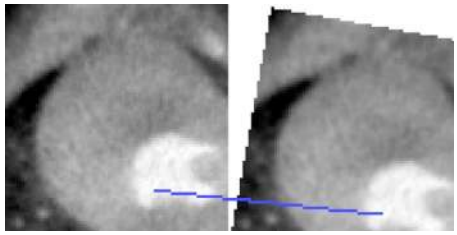

Fig. 4. Line connects an example point pair corresponding to matched n-SIFT features in 4-D CT slices Table VI.

TABLE VI
MEAN FRACTION OF n-SIFT FEATURES MATCHED ACCURATELY (UNDER ERROR IN VOXELS)

Image Type	Transform Type	Matches (< 1.5)	Matches (< 3.0)	Matches (< 7.5)
3D MRI	scale	0.701	0.922	0.991
3D MRI	rotate	0.827	0.897	0.941
3D MRI	rotate+scale	0.641	0.801	0.887
4D CT	rotate	0.850	0.900	0.900

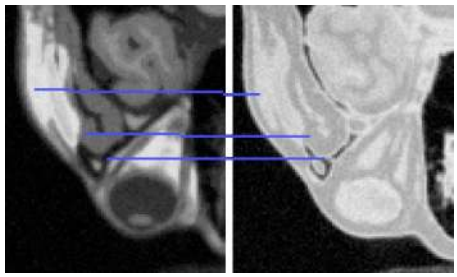

Fig. 5. Lines connect example point pairs corresponding to matched n-SIFT features between Brainweb T1 (slice 32) and PD (slice 36).

2) Reoriented Histogram Feature Evaluation: Table V shows that, for 3-D MRI data, the RGWH results in less accuracy on scaled images than the GWH, although still managing an accuracy of 0.8 at the largest margin of error. However, the accuracy when rotation is considered is greatly improved, with over half the points accurately matched at even the most stringent margin of error. Although the accuracy drops again when we consider both transformations at the same time, it still manages to accurately match over half the points if we consider the largest margin of error. On 4-D CT data, it performs slightly

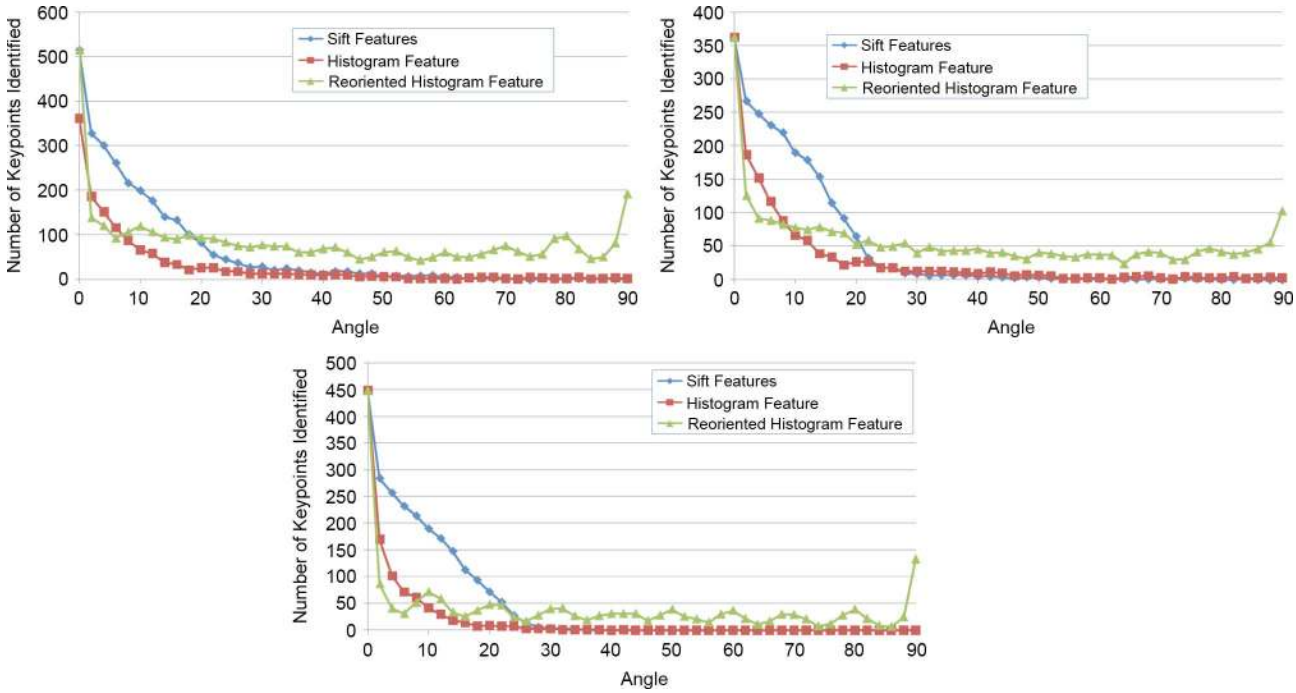

Fig. 6. Number of attempted matches (< 1.5 voxels error) of the original image against the rotated test image for the three 3-D BrainWeb MRI test images.

better than the (GWH), with over 0.75 accuracy at the highest error margin.

3) n-SIFT Feature Evaluation: As seen in Table VI, in all cases the accuracy using the proposed n-SIFT feature was better than the GWH feature, and in all cases, over half of the matches were within 1.5 voxels of error. Over three-quarters of the features were accurately matched at largest margin of error for all tests, with accuracy being over 0.95 if we only look at the scale transformation. At even the most stringent error margin, all tests matched well over half the features accurately. Fig. 4 shows a point from a volume slice at one time point in the original $3\mathrm{D} + \mathrm{time}$ series correctly matched by the algorithm to the corresponding volume slice and time point in the synthetic rotated $3\mathrm{D} + \mathrm{time}$ series.

4) Robustness to Translation: As extrema detection and feature generation both only involve properties local and relative to the extrema point, the number of extrema generated and matching performance was unaffected by translation (Data not shown).

5) Robustness to Rotation: We then looked at the number of matches generated by the various methods, when matching points to the rotated image. As the angle of rotation increases, the ability of all the methods except the RGWH feature should decrease substantially. As we see in Fig. 6, the number of matches generated by the method decreases—in the presence of limited or no good matches, the nonreoriented methods attempt very few (if any) matches. Of note as well is that the number of matches attempted by the reoriented histogram varies in a periodic manner (our analysis of this behavior is presented in Section IV-B6.

We then examined the ability to extract features from rotated test images for angles 0 to 90 degrees at 2 degree increments (Fig. 7). All of the methods extract a large number of keypoints from the original test image; however, even a small amount of rotation significantly decreases the number of discernable keypoints. For the first test image, the GWH feature rapidly deteriorates, with less than 10 keypoints correctly matched after 18 degrees of rotation. The n-SIFT feature reaches this point at 28 degrees, and the RGWH feature matches more than 10 keypoints for all but 2 of the tested angles of rotation. Similar results were obtained for the other two 3-D MRI BrainWeb test images.

We also collected data on the fraction of keypoints matched correctly depending on the amount of rotation of the synthetic test image (Fig. 8). All methods performed were effective at the lowest levels of rotation. The GWH feature rapidly declined to less than 0.6 of identified keypoints matching after only 10 degrees of rotation. As expected, the RGWH feature was the only feature with any reasonable rate of matching after significant (more than 30 degrees) of rotation. The n-SIFT feature, however, proved surprisingly robust despite the lack of rotational invariance, maintaining reasonable levels (significantly more than half) of keypoint matching past 20 degrees of rotation. Also interesting to note is the periodic nature of the accuracy of the response, especially visible in the RGWH feature.

6) Effect of the Number of Histogram Bins: We hypothesize the periodic effect is due to the discrete nature of histogram binning (the parameter b). We note that for the RGWH feature, we use $b = 36$ and so we expect approximately equivalent response at $360/36 = 10$ degree intervals, as was observed. Comparing the fraction of points successfully matched using the RGWH feature on the same image using $b = 18$ shows a similar periodic effect with an extended period of 20 degrees, as expected (Fig. 9). However, other effects also appear to be involved, such as the discrete nature of the image—voxels in the rotated image do not maintain a 1 to 1 correspondence with the original voxels.

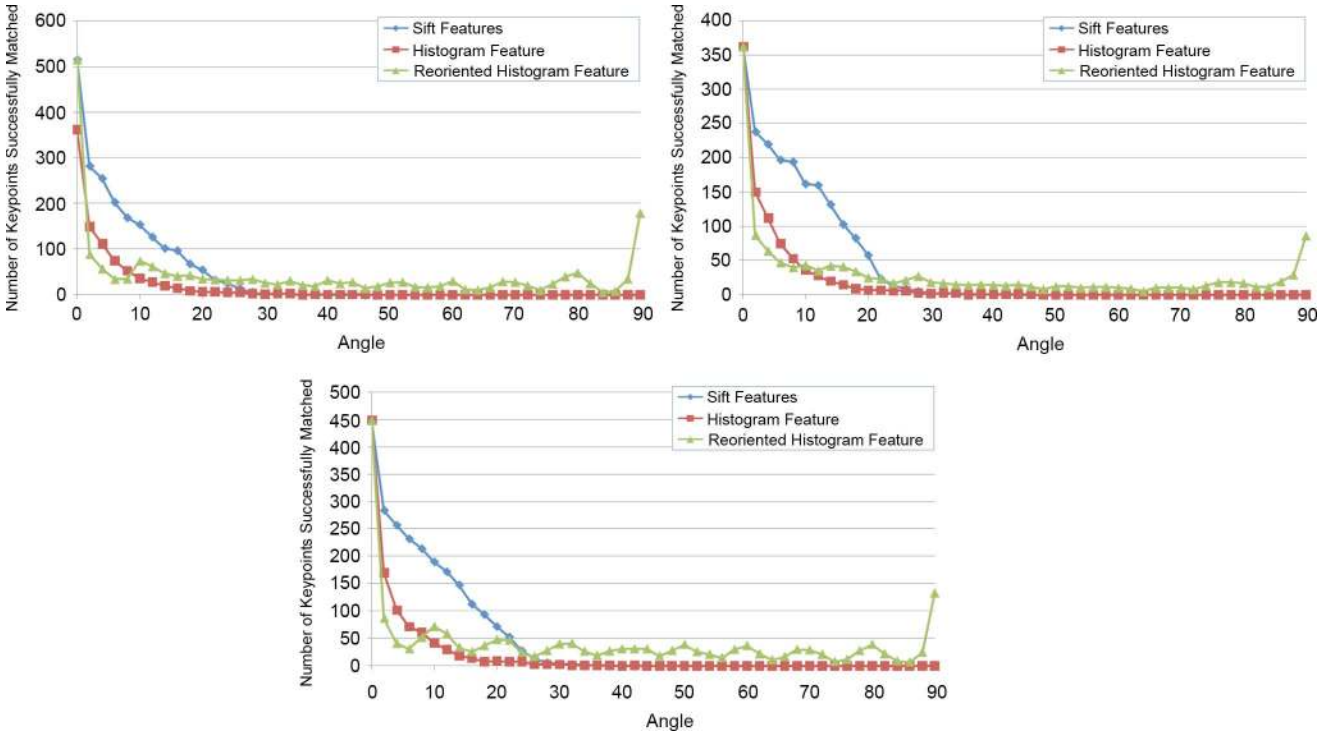

Fig. 7. Number of matches (< 1.5 voxels error) of the original image against the rotated test image for the three 3-D BrainWeb MRI test images.

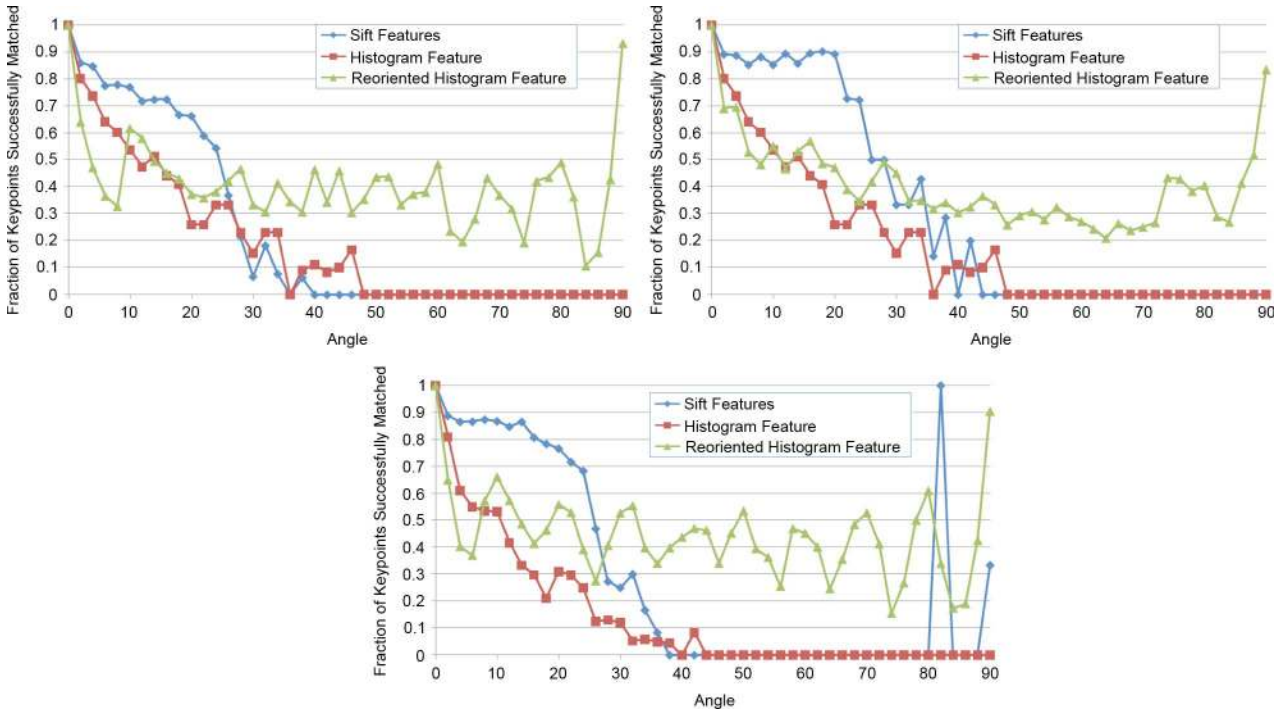

Fig. 8. Fraction of keypoints identified that were successfully matched (< 1.5 voxels error) from the original image against the rotated test image for the three 3-D BrainWeb MRI test images.

7) Robustness to Noise: To evaluate the stability of the features to image noise, we added Gaussian noise [28] with mean 0 and varying σ to the 3-D BrainWeb MRI test images. Increasing σ results in decreasing ability to match points (see Fig. 10). Fig. 11 shows the effect varying sigma (compared to the initial image intensities of 0 to 255). Accurate matching (< 1.5 voxel error) degrades quickly for all the features, with n-SIFT holding a slight advantage most of the time, and the number of feature points matched also drops as the σ increases. Approximate matching (at < 7 voxels errors), however, can remain quite high, especially in the case of n-SIFT, where (for all images), over 90% of points can be matched at this rough threshold to $\sigma = 200$ (although only 2 to 11 points are being matched at this extreme threshold).

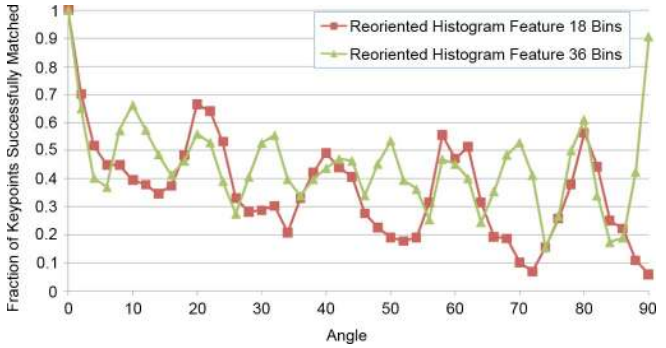

Fig. 9. Comparison of keypoints successfully matched from (< 1.5 voxels error) from the original image against a rotated 3-D BrainWeb MRI test image using the reoriented histogram feature image, using 18 and 36 histogram bins (results for the other two 3-D BrainWeb MRI test images are comparable). Note the periodic response with cycle length of approximately 10 degrees in the case of the 36 bin histogram feature, which becomes a cycle of approximately 20 degrees in the case of the 18 bin histogram feature.

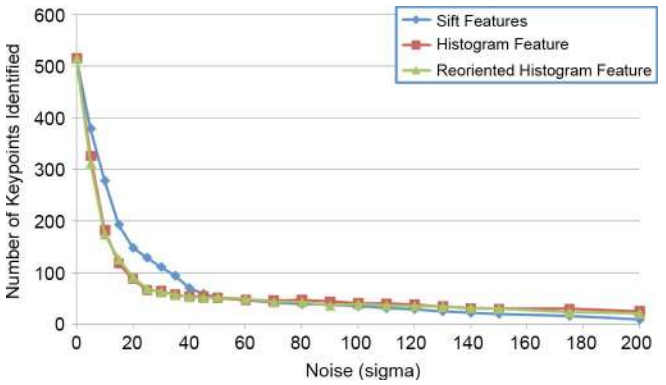

Fig. 10. Number of attempted matches (< 1.5 voxels error) of the original image against Gaussian noised test images for one of the 3-D MRI BrainWeb images (results for other BrainWeb images were similar).

8) Matching Images Between Modalities: Although the features described are intended for matching features captured using the same sensor, generating features on images of differing modalities (but of the same dimensionality) results in feature vectors of the same length. These features can still be compared using method described in Section III-C and we expect successful matches when the local gradients (i.e., the relative response of the various tissues) are comparable. As seen in Table VII where we attempt to match features generated on the BrainWeb test data at different modalities, all featureswere unable to accurately match features from the T1 image to PD or T2 images, but perform fairly well matching from T2 to PD images. However, GWH and n-SIFT were able to inaccurately match over half the points from T1 to PD (n-SIFT was unable to find many points to match from T1 to T2) (see Fig. 5).

V. Discussion

We propose three SIFT-like features for matching points in images—a global weighted histogram (GWH) feature, a reoriented global weighted histogram RGWH feature and the n-dimensional Scale-Invariant Feature Transform (n-SIFT) feature.

We show the GWH feature is sufficient for robustness to scaling, however, cannot cope with 10° rotation. The RGWH feature reorients the image based on local gradient vectors before computing the feature vector, improving robustness to rotation, as shown by the improved accuracy with rotation. However, the decreased performance when we consider the scaling transform shows that reorientation sacrifices the distinctiveness of the histogram, as the largest bin is always reoriented to be in the same position in the feature.

n-SIFT was the most successful of all the features tested. It incorporates the discriminatory power of the gradient-based features, but further subdivides the image into regions, each with a separate histogram, relative to the feature location. Although not reoriented, it is robust against minor rotational change, as it employs only 8 bins to summarize each of the hyperspherical coordinates, with each bin covering a 45° range.

TABLE VII
Fraction of Features Matched Accurately (Under Error in Voxels) Across Different Modalities Using Different Features

Feature Type	Image Modalities	Points Matched	Matches (< 1.5)	Matches (< 3.0)	Matches (< 7.5)
GWH	T1 and PD	64	0.125	0.25	0.609
GWH	T1 and T2	51	0.118	0.215	0.549
GWH	T2 and PD	87	0.345	0.529	0.770
RGWH	T1 and PD	65	0.092	0.169	0.462
RGWH	T1 and T2	49	0.102	0.163	0.306
RGWH	T2 and PD	79	0.354	0.519	0.734
n-SIFT	T1 and PD	12	0	0.167	0.583
n-SIFT	T1 and T2	3	0	0	0
n-SIFT	T2 and PD	54	0.667	0.889	0.944

A. Performance Considerations

The prototype system used in this paper was implemented in C++, using the Insight Toolkit (ITK) and provided as an open source publication [24], which includes tests to recreate part of our results. Image sizes were cropped for testing of the many combinations of images, features and synthetic test conditions.

The computation time of the feature extraction methods described scales linearly with the size of the images involved. However, as the dimensionality of the image increases, the image size can also increase dramatically. The current implementation is single-threaded; however, it should be possible adapt strategies used for optimizing 2-D SIFT implementations, such as the multicore implementation of 2-D SIFT [29], [30], and the GPU-based implementation [31].

B. Future Work

Even small error can have critical impact in medical applications. However, not all techniques used with SIFT features in 2-D images [5] have been implemented. Localization of features could be improved by removing difficult-to-localize edge-like extrema via testing ratios of principal curvatures, as well as interpolating positions of the extrema using the Taylor expansion of the doG scale space function. Images could also be upscaled by a factor of 2 before running n-SIFT to sample higher frequency information. For the feature vector itself, reorienting of the feature vectors and trilinear interpolation of the samples

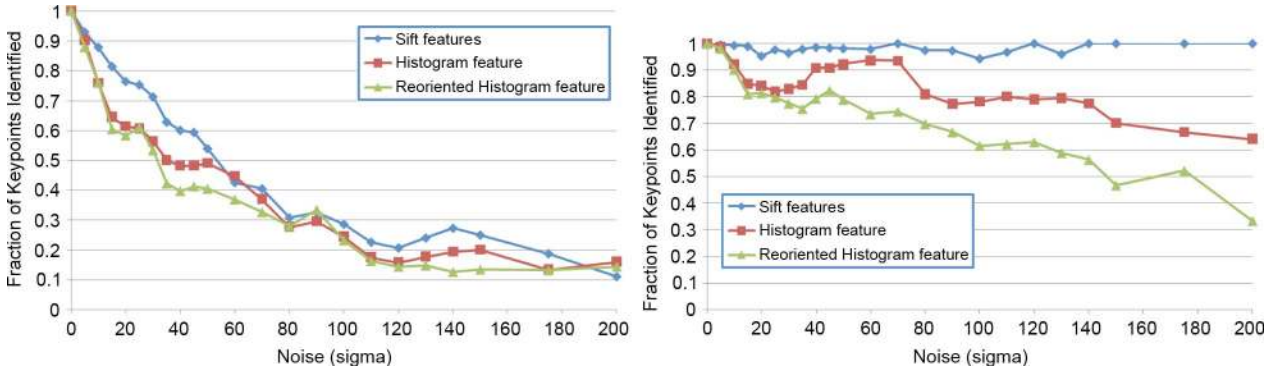

Fig. 11. Fraction of keypoints identified that were successfully matched at < 1.5 voxels error (left) and < 7 voxels error (right) from the original image against Gaussian noised image for one of the three 3-D BrainWeb MRI test images.

could also be implemented. Additionally, the n-SIFT histogram features could be normalized, potentially improving matching, especially between modalities, as in Chen and Tian [19].

All the parameters used were left at their default values from [5] and [25] for the tests. More appropriate values may exist for higher-dimensional images and for particular applications. n-SIFT could also initialize other intensity-based registration methods, or potentially adapted for nonscalar vector or tensor fields obtained from Diffusion Tensor MRI.

As long as extrema can be localized in approximately the location and the local gradient near these extrema are comparable, the techniques described here remain effective. This can allow for matching even under large-scale image deformations, as long as the local deformations are not too extreme. Already described here is a method to compensate for significant rotational change in the local gradients. A future extension, therefore, would be to detect and compensate for other deformations, further improving the robustness of the features.

We do not claim the features described here as a replacement for existing image registration approaches. Rather, the techniques described could be used to provide corresponding landmarks that initialize or provide hard/soft constraints for other image registration approaches. Comparing the performance of these methods in combination with various registration approaches may reveal other avenues to improve the features.

Another application of these features would be for image retrieval, matching the features generated from an image against the features from a library of images. Modified SIFT features have been applied to this problem in the 2-D case [32]; however, one extension would be to apply this to libraries of higher-dimensional images.

C. Conclusion

n-SIFT extends SIFT features to images of arbitrary number of dimensions. We evaluate a difference of Gaussian keypoint detector, a global histogram of gradient feature vectors and reorientation of this vector in the direction of the most significant gradient, and the n-SIFT features. Our results indicate these SIFT-like features can be matched efficiently in 3-D and 4-D images, making them a potentially effective tool to find corresponding landmarks in related images. The implementation has been provided as open source for the community to use, test and extend, and has been downloaded by over 250 users at our website from November 2007 to May 2009, in addition to 2267 views and 1134 downloads at the *Insight Journal* (July 16, 2009).

REFERENCES

[1] Z. Xue, D. Shen, and C. Davatzikos, "Determining correspondence in 3-d mr brain images using attribute vectors as morphological signatures of voxels," *IEEE Trans. Med. Imag.*, vol. 23, no. 10, pp. 1276–1291, Oct. 2004.
[2] T. F. Cootes, D. Cooper, C. J. Taylor, and J. Graham, "Active shape models—their training and application," *CVIU*, vol. 61, no. 1, pp. 38–59, 2004.
[3] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," *IEEE Pattern Anal. Mach. Intell.*, vol. 23, no. 6, pp. 681–685, Jun. 2001.
[4] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. IEEE Int. Conf. Computer Vision*, 1999, vol. 2, p. 1150.
[5] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
[6] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 8, pp. 679–714, 1986.
[7] C. Harris and M. Stephens, "A combined corner and edge detector," in *Proc. 4th Alvey Vision Conf.*, 1988, pp. 146–151.
[8] J. Shi and C. Tomasi, "Good features to track," presented at the 9th IEEE Conf. Computer Vision and Pattern Recognition, 1994.
[9] S. M. Smith and J. M. Brady, "Susan–A new approach to low level image processing," *Int. J. Comput. Vis.*, vol. 23, pp. 45–78, 1997.
[10] M. Trajkovic and M. Hedley, "Fast corner detection," *Image Vis. Comput.*, vol. 16, pp. 75–87, 1998.
[11] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide baseline stereo from maximally stable extremum regions," in *Proc. Brit. Machine Vision Conf.*, 2002, pp. 384–393.
[12] T. Lindeberg, "Detecting salient blob-like image structures and their scales with a scale-space primal sketch: A method for focus-of-attention," *Int. J. Comput. Vis.*, vol. 11, pp. 283–318, 1993.
[13] T. Lindeberg, "Feature detection with automatic scale selection," *Int. J. Comput. Vis.*, vol. 30, pp. 77–116, 1998.
[14] R. G. Stevens and S. A. Beaman, "The use of difference of gaussian image filtering to assess objectively the correlation between breast vascularity and breast cancer," *Phys. Med. Biol.*, vol. 33, no. 12, pp. 1417–1431, 1988.
[15] Y. Ke and R. Sukthankar, "PCA-SIFT: A more distinctive representation for local image descriptors," *IEEE Comput. Vis. Pattern Recognit.*, vol. 2, pp. 506–513, 2004.
[16] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," in *Proc. IEEE CVPR*, 2003, vol. 2, pp. 257–263.
[17] M. Moradi and P. Abolmaesumi, "Medical image registration based on distinctive image features from scale-invariant (SIFT) keypoints," *CARS*, p. 1292, 2005.
[18] M. Moradi, P. Abolmaesoumi, and P. Mousavi, "Deformable registration using scale space keypoints," *Proc. SPIE Med. Imag.*, vol. 6144, pp. 6144-2G1–6144-2G8, 2006.

[19] J. Chen and J. Tian, "Rapid multi-modality preregistration based on SIFT descriptor," in *Proc. 28th Annu. Int. Conf. IEEE Engineering in Medicine and Biology Society*, 2006, pp. 1437–1440, DOI: 10.1109/iembs.2006.260599.

[20] B. Qiu, "A refined SVM applied in medical image annotation," in *Proc. Evaluation Multilingual and Multi-Modal Information Retrieval*, 2007, vol. 4730, pp. 690–693, DOI: 10.1007/978-3-540-74999-8.

[21] A. Franz, I. Carlsen, and S. Renisch, "An adaptive irregular grid approach using SIFT features for elastic medical image registration," *Bildverarbeitung für die Medizin 2006*, pp. 201–205, DOI: 10.1007/3-540-32137-3_41.

[22] M. Urschler, J. Bauer, H. Ditt, and H. Bischof, "Sift and shape context for feature-based nonlinear registration of thorac ic ct images," in *Proc. Computer Vision Approaches to Medical Image Analysis*, 2006, vol. 4241, pp. 73–84, DOI: 10.1007/11889762_7.

[23] W. Cheung and G. Hamarneh, "n-SIFT: n-dimensional scale invariant feature transform for matching medical images," in *Proc. 4th IEEE Int. Symp. Biomedical Imaging: Biomedical Imaging: From Nano to Macro*, 2007, pp. 720–723, DOI: 10.1109/ISBI.2007.356953.

[24] W. Cheung and G. Hamarneh, "Scale invariant feature transform for n-dimensional images (n-SIFT)," *The Insight J.*, vol. 2007, Jul.–Dec. 2007.

[25] S. Nowozin, *Autopano-SIFT ver. 2.4.* Accessed July 2009. [Online]. Available: http://user.cs.tu-berlin.de/~nowozin/autopano-sift/

[26] C. A. Cocosco, V. Kollokian, R. K.-S. Kwan, and A. C. Evans, "Brainweb: Online interface to a 3-D MRI simulated brain database," in *Proc. HBM NeuroImage*, 1997, vol. 5, no. 4, p. S425.

[27] D. R. Holmes, III, E. L. Workman, and R. A. Robb, "The NLM-Mayo image collection: Common access to uncommon data," presented at the MICCAI Open-Source Workshop, 2005.

[28] G. Baker, ITK Noise Generators 2005, Accessed July 2009. [Online]. Available: http://www.cs.mu.oz.au/~gavinb/projects/itk.php#noise

[29] H. Feng, E. Li, Y. Chen, and Y. Zhang, "Parallelization and characterization of SIFT on multi-core systems," in *Proc. IEEE Int. Symp. Workload Characterization*, Sep. 2008, pp. 14–23.

[30] Q. Zhang, Y. Chen, Y. Zhang, and Y. Xu, "SIFT imlpementation and optimization for multi-core systems," in *Proc. IEEE Int. Symp. Parallel and Distributed Processing*, Apr. 2008, pp. 1–8.

[31] S. Sinha, J.-M. Frahm, M. Pollefeys, and Y. Genc, "GPU-based video features tracking and matching," presented at the Workshop on Edge Computing Using Parallel and Distributed Processing, Chapel Hill, NC, May 2006.

[32] T. Tommasi, F. Orabona, and B. Caputo, "Discriminative cue integration for medical image annotation," *Pattern Recognit. Lett.*, vol. 29, no. 15, pp. 1996–2002, 2008.

Warren Cheung received the B.Sc. degree in Microbiology and Computer Science from the Faculty of Science, University of British Columbia (UBC), Vancouver, BC, Canada, in 2002, and the M.Sc. degree from the Department of Computer Science, UBC, in 2005. He is currently pursuing the Ph.D. degree in bioinformatics at the University of British Columbia under the Canadian Institutes for Health Research/Michael Smith Foundation for Health Research Strategic Program in Bioinformatics.

His research interests include algorithms and computational analysis relevant to biomedical problems.

Ghassan Hamarneh received the B.Sc. degree from the Department of Electrical Engineering, Jordan University, Jordan, in 1995, and the M.Sc. degree with distinction in digital communications from the Chalmers University of Technology, Sweden, in 1997. He completed his doctoral studies (2001) in the Department of Signals and Systems, Chalmers University of Technology, and as a Predoctoral Research Fellow (2000–2001) in the Department of Computer Science, University of Toronto, Toronto, ON, Canada.

Before joining the School of Computing Science, Simon Fraser University, Burnaby, BC, Canada (as an Assistant Professor in 2003 and as an Associate Professor in 2009), he was a Postdoctoral Fellow at the Hospital for Sick Children, Mouse Imaging Centre, and the Department of Medical Biophysics, University of Toronto, Toronto, ON, Canada (2001–2003). He is the Co-Founder (2003) and Co-Director of the Medical Image Analysis Laboratory (MIAL), Simon Fraser University. His main research interests are medical image processing, segmentation and registration, and anatomical shape modeling and analysis.