N-SIFT: N-DIMENSIONAL SCALE INVARIANT FEATURE TRANSFORM FOR MATCHING MEDICAL IMAGES

Warren Cheung

Bioinformatics
University of British Columbia

Ghassan Hamarneh

Medical Image Analysis Lab
Simon Fraser University

ABSTRACT

We present a fully automated multimodal medical image matching technique. Our method extends the concepts used in the computer vision SIFT technique for extracting and matching distinctive scale invariant features in 2D scalar images to scalar images of arbitrary dimensionality. This extension involves using hyperspherical coordinates for gradients and multidimensional histograms to create the feature vectors. These features were successfully applied to determine accurate feature point correspondence between pairs of medical images (3D) and dynamic volumetric data (3D+time).

Index terms — *Biomedical image processing, Image matching, Feature extraction*

1. 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, 2, 3]. The Scale-Invariant Feature Transform (SIFT) produces stable features in two-dimensional images[4, 5]. We extend 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 image gradient, weighted depending on 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 a standard n-SIFT 4^n histogram feature vector. We use nD gradient vectors with $n-1$ directions and a magnitude, summarised using a weighted multidimensional histogram. As well, the 128D SIFT feature vector has been generalised to a 2^{5n-3}D feature vector.

DoG image filtering has been previously used to successfully identify features in breast thermograms for early detection of breast cancer[6], indicating that extrema in doG may also indicate salient features. SIFT features and its extensions have been previously applied for object recognition as an easily computed scale-space kernel approximating the scale-normalised Laplacian of Gaussian[4, 5, 7] and have been shown to be most resistant against common image deformations[8]. It has also been evaluated on 2D medical images[9], showing stability under arbitrary affine transformations.

In this paper, we present a detailed description of the proposed extension (Section 2), then evaluate the features using 3D and 4D medical image data (Section 3) and conclude by discussing the results (Section 4).

2. METHODS

2.1. Feature Localisation

To achieve invariance to image scaling, we use a multilevel image pyramid extending that of 2D SIFT[4, 5]. At each successive level, the first image (represented by white boxes in Fig. 1) is a linearly interpolated downsampled version of the Gaussian smoothed image at scale σ in the previous level. This yields a level for the original image, a level of half-sized images, quarter-sized, and so on. At each level, we sample the doG space. First, we generate a series of Gaussian blurred images from the first image using sigma $= \sigma, k\sigma, k^2\sigma$. A doG image at scale $k^j\sigma$ is generated from the difference of neighbouring blurred images at scales $k^j\sigma$ and $k^{j+1}\sigma$. To test s doG image scales in a pyramid level, we set $k^s = 2$.

For each pyramid level, we locate extrema in our approximation of doG space. Each voxel of a doG image (scale $k^j\sigma$) is compared against the neighbouring voxels immediately adjacent, including diagonal directions, its corresponding voxel in the scale above ($k^{j+1}\sigma$) and all the neighbours to that corresponding voxel, and the corresponding voxel in the scale below ($k^{j-1}\sigma$) and all the neighbours. For an nD image, there are a total of $3^{n+1}-1$ neighbours to be checked. (Fig. 2).

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 are needed. Finally, only extrema in the doG image having magnitude greater than a threshold T_{doG} are considered feature points.

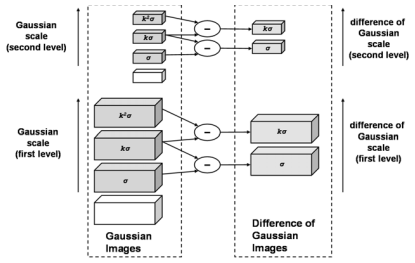

Fig. 1. Image pyramid — Each level is a downsampled image, convolved with Gaussian scales σ, $k\sigma$, $k^2\sigma$ and so on, and doG images from adjacent scales.

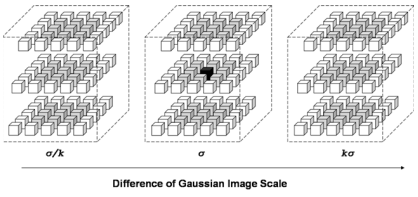

Fig. 2. For 3D images, an extremum (in black) is the maximum or minimum of the neighbours (shaded) at the same scale, and the corresponding voxels in the scale above and below.

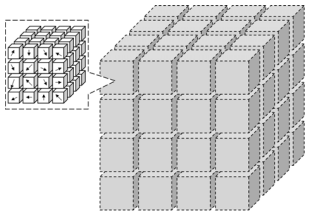

Fig. 3. In 3D, each of the $4 \times 4 \times 4$ regions (dashed and shaded) summarises the gradients at $4 \times 4 \times 4$ voxel locations (solid and white).

2.2. Feature Generation

The *global histogram feature* simplifies the SIFT feature to a single histogram, summarising the gradients near the located feature point. This is effectively the histogram used by SIFT to reorient the feature vector. In nD, we represent the direction of the gradient via $n-1$ hyperspherical coordinates and use a $n-1$ dimensional histogram with b bins spanning 2π for each orientation, for a total of b^{n-1} bins. For each voxel in the image, the magnitude of the gradient is weighted by a Gaussian centred on the feature position. This value is added to the corresponding bin for the gradient orientation.

The *reoriented global histogram feature* normalises the direction of the gradients with respect to the highest peak in the histogram, increasing robustness to changes in orientation. The index for the bin for this peak is first converted to its average orientation direction. A new histogram is then created by taking each bin in the histogram, converting its bin index to hyperspherical coordinates, rotating it by subtracting the peak orientation vector, and then converting this vector to the bin index of the new histogram where we store the contents of the bin.

The *n-SIFT feature* summarises a 16^n hypercubic voxel region around the feature position, and divides this into 4^n subregions, each using a 8^{n-1} bin histogram to summarise the gradients of the voxels in the subregion, resulting in a $4^n 8^{n-1} = 2^{5n-3}$-dimensional feature vector (Fig. 3). The n-SIFT feature implemented is analogous to the 2D SIFT feature[5], except it does not reorient the feature vectors or apply trilinear interpolation of the samples. The histograms summarise the gradient direction, weighted by a Gaussian centred at the feature position.

2.3. Feature Matching

To match histograms generated by the three types of features, we consider the histogram(s) of b bins of a feature as a b-dimensional vector, and compare the l^2 distance between a feature in one image against every feature 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 $\frac{d(v,u)}{d(v,u')}$, 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.

3. RESULTS AND VALIDATION

We applied these features to 3D human Magnetic Resonance Imaging (MRI) brain scans and 4D 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`[10]. 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 global histogram feature, we use $b = 36$ and Gaussian $\sigma = 8.0$ for weighting, matching the values used for reorienting the SIFT feature vector. For n-SIFT, we use $b = 8$ and $\sigma = 4.0$. The matching threshold T_m was 0.8 . n-SIFT was implemented in C++ using ITK.

Three 1mm isotropic BrainWeb simulated brain 3D MRI data (cropped size $90 \times 108 \times 90$)[1][11] were used. 4D CT images of in-vivo canine heart from the Mayo Dynamic Spatial Reconstructor, 0.925mm isotropic $\times 18$ images per cardiac cycle (cropped size $78 \times 80 \times 88$ voxels $\times 9$ time points)[2][12], were also used. Extrema were searched through three octaves. The images were transformed using linear interpolation. Tests were run on an Intel Pentium M 1.7 Ghz machine with 1Gb of memory. Runtime on all sample 3D images was less than 19 CPU minutes for over 874 thousand voxels.

3.1. Evaluating the DoG Extrema Extraction

For the 3D images, extracting the extrema took 100–102 CPU seconds. Generating and matching feature vectors were much

[1] `ftp://public.kitware.com/pub/itk/Data/BrainWeb`
[2] `http://nova.nlm.nih.gov/Mayo/NLMDATA/Dog/`

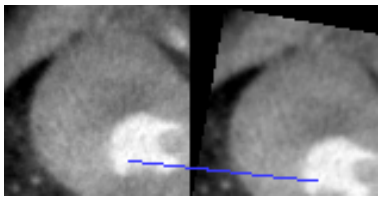

Fig. 4. Line connects an example point pair corresponding to matched n-SIFT features in 4D CT slices (See Table 5).

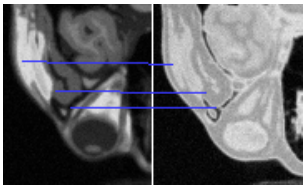

Fig. 5. Lines connect example point pairs corresponding to matched n-SIFT features between Brainweb T1 and PD slices.

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

Table 1. Number of Extrema Localised for the Brainweb 3D image, and the same image downscaled by 0.8. The matches generated by the global histogram, reoriented global histogram and n-SIFT features are compared.

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 2. Stability of Extrema for Transformed 3D MRI images. Matches are the mean of the fraction correctly matched extrema, under various error criteria in voxels. The transformations are scaling by a factor of 0.8, rotating by $10°$, or both scaling and rotating.

more computationally intensive parts. The total runtime with these steps was 530–740 CPU sec for global and global reoriented histogram features, with reorientation causing no more than 10 CPU sec difference. n-SIFT is the quickest of the features to run, taking a total of 330-415 CPU sec, although reorienting the n-SIFT feature would negate this by requiring generation of the global histogram.

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 were compared. Each extremum located in the scaled image is then matched with the closest corresponding extremum in the original image, and the margin of error computed. Table 1 shows half to two-thirds of global histogram 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 2 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.

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

Table 3. Mean fraction of global histogram-based features matched accurately (under error in voxels).

3.2. 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.

Table 3 shows the global histogram feature, for 3D 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 4D CT data.

Table 4 shows that, for 3D MRI data, reorienting the global histogram results in less accuracy on scaled images, 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 4D CT data, it performs slightly better than the global histogram, with over 0.75 accuracy at the highest error margin.

As seen in Table 5, in all cases the accuracy using the n-SIFT feature was better than the simple histogram 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

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

Table 4. Mean fraction of reoriented histogram-based 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

Table 5. Mean fraction of n-SIFT features matched accurately (under error in voxels).

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 three points from a volume slice at one time point in the original time series correctly matched by the algorithm to three corresponding points on the rotated test time series.

4. CONCLUSIONS

We propose three SIFT-like features for matching points in images — a global histogram feature, a reoriented global histogram feature and the n-SIFT feature. Fig. 5 shows n-SIFT matching points between brain MRI of differing modalities.

The global histogram-based feature is sufficient for robustness to scaling, however, cannot cope with 10° rotation. Reorienting the image before computing the global histogram feature makes the feature more robust to rotation, as shown by the improved accuracy. However, the decreased performance when we consider the scaling transform shows this 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 this rotational change, as it employs only 8 bins to summarise each of the hyperspherical coordinates, with each bin covering a 45° range.

Even small error can have critical impact in medical applications. However, not all techniques used with SIFT features in 2D images[5] have been implemented. Localisation of features can be improved by removing edge-like extrema via testing ratios of principal curvatures, as well as interpolating positions of the extrema in doG space. Images could also be upscaled by a factor of 2 before running n-SIFT to sample higher frequency information.

Note as well that all the parameters used were left at their default values for the tests. More appropriate values may exist for higher-dimensional images and for particular applications. n-SIFT could also initialise other intensity-based registration methods, or potentially adapted for non-scalar vector or tensor fields obtained from Diffusion Tensor MRI.

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. A results indicate these SIFT-like features can be matched efficiently in 3D and 4D images, making them a potentially effective tool to find corresponding landmarks in related images.

5. 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 TMI*, vol. 23, no. 10, pp. 1276–1291, 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 PAMI*, vol. 23, no. 6, pp. 681–685, 2001.

[4] D. G. Lowe, "Object recognition from local scale-invariant features," in *IEEE ICCV*, 1999, vol. 2, p. 1150.

[5] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[6] 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.

[7] Y. Ke and R. Sukthankar, "PCA-SIFT: A more distinctive representation for local image descriptors.," in *IEEE CVPR*, 2004, vol. 2, pp. 506–513.

[8] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," in *IEEE CVPR*, 2003, vol. 2, pp. 257–263.

[9] M. Moradi and P. Abolmaesumi, "Medical image registration based on distinctive image features from scale-invariant (SIFT) keypoints," in *CARS*, 2005, p. 1292.

[10] S. Nowozin, "Autopano-SIFT ver. 2.4," 2005.

[11] C. A. Cocosco, V. Kollokian, R. K.-S. Kwan, and A. C. Evans, "Brainweb: Online interface to a 3d mri simulated brain database," *NeuroImage — Proc. of HBM*, vol. 5, no. 4, pp. S425, 1997.

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