Keypoint Recognition using Two-Stage Randomized Trees

Shoichi Shimizu
Advanced Technology R&D Center, Mitsubishi Electric Corporation
8-1-1, Tsukaguchi-Honmachi, Amagasaki, Hyogo 661-8661, Japan
Shimizu.Shoichi@ab.MistubishiElectric.co.jp

Hironobu Fujiyoshi
Dept. of Computer Science, Chubu University
1200 Matsumoto, Kasugai, Aichi 487-8501 Japan
hf@cs.chubu.ac.jp

Abstract

This paper proposes a high-precision, high-speed keypoint matching method using a two-stage Randomized Trees. The keypoint classification method uses the conventional Randomized Trees to enable high-precision, real-time keypoint matching. But the wide variety of view transformations for templates expressed by Randomized Trees make high-precision keypoint classification for all transformations difficult with a single Randomized Trees. To resolve this problem, proposed method classifies the template view transformations during the first stage. Then during the second stage, it classifies the keypoints using the Randomized Trees corresponding to each of the view transformations classified during the first stage. For images in which the viewpoint of the object is rotated by 70 degree, evaluation testing demonstrated that proposed method is 88.4% more precise than SIFT, and 63.4% more precise than the conventional Randomized Trees. We have also shown that the proposed method supports real-time keypoint matching at a speed of 12 fps.

1 Introduction

Technology for automatic recognition of specific objects in images holds promise for implementation in a variety of fields and is an important research topic in the field of computer vision. Implementation of such specific object recognition requires a recognition method that is robust against view changes such as image rotation, changes in scale, changes in illumination, and changes in viewpoint. Moreover, real-time processing is also important.

Conventional methods that use local features for corresponding point matching can be divided into two types, those that use high-performance local features and those that introduce a training algorithm. The former type is typified by Scale Invariant Feature Transform (SIFT) [1]. SIFT is robust against image rotation, changes in scale, and changes in illumination, and so is capable of highly accurate matching. In recent years, PCA-SIFT [2], which increases the descriptive power of SIFT, GLOH [3], and Shape Context [4] have been proposed to achieve higher matching accuracy. However, these SIFT based approaches suffer from the problem of high computational cost. Although faster versions of SIFT (SURF [5] and Fast Approximated SIFT [6]) have been proposed, real-time processing remains difficult at this time.

On the other hand, a method that uses a training algorithm to train Randomized Trees (RTs) for keypoint classification has been proposed [7]. Reference [7] applies affine transforms to generate training images that represent various pseudo view changes from a single template image. Those images are then used for RTs training [8], allowing keypoint classification that is robust to view changes. The RTs technique implements corresponding point search by decision tree traversal and is capable of high-speed classification of keypoints. In recent years, this method has been developed further, and there are reports that it can even run on low-memory mobile devices [9, 10]. However, methods based on reference [8] have the problem of low matching accuracy when there are large view changes in the image. One cause of that problem is that there are various kinds of view changes in the template represented by RTs, so a single RT cannot easily achieve highly accurate keypoint classification with respect to all of the view changes.

To solve that problem, we propose here a keypoint classification method that uses two-stage Randomized Trees. The proposed method classifies the viewpoints of the input image in the first stage; in the second stage, keypoint classification is performed using the RTs trained with image viewpoints that are near those classified in the first stage. Thus, because a RT that has been trained on images visually close to the input image can be used for the keypoint classification, improved keypoint classification can be expected.

2 Proposed method: corresponding point matching with two-stage Randomized Tree

The proposed method deals with changes in template viewpoint and keypoint classification by training two-stage RTs. In the first stage, the input image viewpoints are classified. Viewpoint classes are groups of the many different viewpoints of a training image that are clustered around K representative viewpoints. In the second stage, keypoint classification is done using the RTs trained with the training images that belong the viewpoint classes identified in the first stage. In this way, the keypoints can be classified with RTs trained with images that have viewpoints close to the input image. The processing flow for the proposed method is shown in Fig. 1. First, training images that represent many different viewpoints of the template are generated and viewpoint clustering is done as preprocessing. Next, the two-stage RTs are trained.

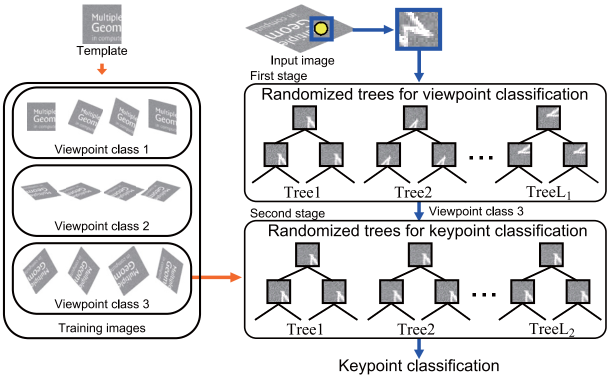
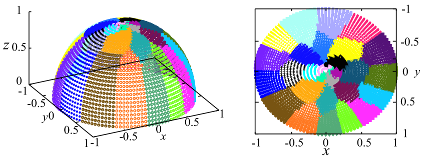


Figure 1. Processing flow for the proposed method.



(a) Side view

(b) Top view

Figure 2. Spherical display of viewpoint clustering results.

2.1 Generation of training images

The training image generation and viewpoint clustering are described in detail below.

2.1.1 Three-dimensional rotation training image

In reference [7], the affine transform parameters for generating the training images are selected randomly, which introduces the problems of viewpoint bias and inability to represent rotation in three dimensions. To solve those problems, the proposed method uses Euler angles to represent rotation in three dimensions when generating training images.Let Eq. (1) be $\mathbf{A}$, the affine transform matrix using the Euler angles in a 2-D coordinate system.

$$\mathbf{A} = \begin{bmatrix} \cos(\psi) & -\sin(\psi) \\ \sin(\psi) & \cos(\psi) \end{bmatrix} \begin{bmatrix} \cos(\theta) & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \cos(\phi) & -\sin(\phi) \\ \sin(\phi) & \cos(\phi) \end{bmatrix} \quad (1)$$

Here, $\mathbf{A}$ is a 2×2 matrix for transformation in a 2-D coordinate system. The viewpoint bias problem is solved by generating the training images from the affine transform matrix $\mathbf{A}$, whose Euler angle rotation parameters ψ, θ and ϕ are set to equal intervals. In the research reported here, the rotation ranges for the parameters are $\phi \in [0°, 90°]$, $\theta \in [0°, 90°]$, and $\psi \in [0°, 360°]$, and the interval for ϕ, θ, and ψ is 5°, and 23,328 training images are generated from one template image.

2.1.2 Viewpoint clustering

Next, viewpoint clustering is done. When clustering by the Euler angle X, Y, and Z axis rotation parameters, the periodicity in the rotation cannot be represented. Therefore, the proposed method clusters the viewpoints by the k-means clustering, using the generated patch images as features. Thus, even for image rotations of 0° or 359°, clustering of images of close viewpoints is possible. For each training image, a series of linked 32 x 32 patch images centered on the keypoints is created. The patch image series is projected into the intensity feature space and clustered by the k-means clustering. The clustering results are presented in Fig. 2 for the number of viewpoint classes $K = 30$, with clusters represented by color coding. From Fig. 2, we see that images whose viewpoints are close can be clustered.

2.2 Two-stage Randomized Trees training

The proposed method deals with changes in template viewpoint and the keypoint classification problem by training two-stage RTs. In the first stage, viewpoint class frequencies are learned by RTs using the relation of patch intensity magnitudes. In the second stage, RTs are trained for each viewpoint class. Accordingly, RTs are created for each of the viewpoint classes indicated by a color in Fig. 2. The two-stage RTs training method is described in detail below.

2.2.1 First stage: Training the viewpoint classification Randomized Trees

The decision tree set $T1 = \{T1_1, \cdots, T1_{L_1}\}$ for classifying the input image viewpoints is trained. L_1 is the number of decision trees. Decision tree set T_1 is trained by dividing the patches into L_1 subsets. Node branching is determined by the intensity magnitude relationship of the keypoint patches in the same way as for reference [7].

$$C_2(\mathbf{m}_1, \mathbf{m}_2) = \begin{cases} \texttt{L If } I_\sigma(\mathbf{P}, \mathbf{m}_1) \leq I_\sigma(\mathbf{P}, \mathbf{m}_2) \\ \texttt{R otherwise} \end{cases} \quad (2)$$

The L and R indicate the left and right child nodes. $I_\sigma(\mathbf{P}, \mathbf{m})$ is the intensity of pixel $\mathbf{m}$ in patch $\mathbf{P}$. Then, the viewpoint class probability distributions of the leaf nodes are obtained. It is thus possible to classify the viewpoints using the probability distribution of the leaf node arrived at during classification. In the research reported here, the number of patches handled in the first stage of RTs is 9.33 million when the number of training images is m=23,328 and the number of keypoint classes is c=400.

2.2.2 Second stage: Training Randomized Trees for keypoint classification

The decision tree sets for keypoint classification are trained for each viewpoint class. The second stage decision tree set comprises the decision tree set of K viewpoint classes, $T2 = \{T2_1, \cdots, T2_K\}$. The decision tree set for viewpoint class $k (k \in K)$, $T2_k = \{T2_{k,1}, \cdots, T2_{k,L_2}\}$. is trained by dividing the patches that belong to viewpoint class k into L_2 subsets. The node branching is determined by the relationship of the intensity magnitudes of the keypoint patches in the same way as Eq. (2). Then, the probability distributions of the leaf node keypoints classes are obtained. In the research reported here, the number of patches handled by RTs of the second stage $T2_k$ is 310,000 when the number of training images is $m = 23,328$, the number of keypoints classes is $c = 400$, and the number of viewpoint classes is $K = 30$.

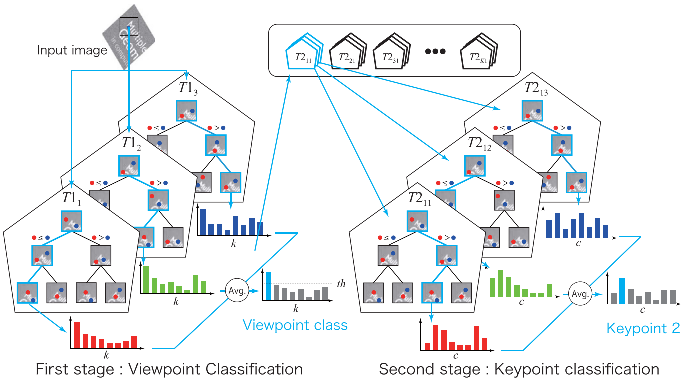


Figure 3. Keypoint classification with two-stage decision trees.

2.3 Keypoint classification using two-stage Randomized Trees

The flow of keypoint classification is shown in Fig. 3. Keypoints are extracted in the same way as reference [7]. Viewpoint class k of the input image is classified with decision tree set $T1$ from the first stage. Next, the keypoints are classified with decision tree set $T2_K$ from the second stage decision tree set $T2$, which corresponds to the viewpoint class k. classified in the first stage.

2.3.1 First stage: Viewpoint class classification

The classification of viewpoint class k involves obtaining the mean of the probability distribution of the leaf node in the decision tree $T1 = \{T1_1, \cdots, T1_{L1}\}$ at which input patch P arrived, $P_{\eta(T1_l,\mathbf{P})}(Y(\mathbf{P}) = k)$, using all of the keypoints on the template as indicated in Eq. (3), and is extracted as viewpoint class k if it exceeds the threshold value th.

$$G(k) = \begin{cases} \texttt{1 If } \frac{1}{L_1}\sum_{l=1}^{L_1} P_{\eta(T1_l,\mathbf{P})}(Y(\mathbf{P})=k)>th \\ \texttt{0 otherwise} \end{cases} \qquad (3)$$

2.3.2 Second stage: Keypoint classification

In the first stage viewpoint class classification, there are cases in which there are multiple classes for which $G(k) = 1$. Thus, for keypoint classification we obtain the mean of the leaf node probability distribution $P_\eta(T2_{kl}, \mathbf{P})(Y(\mathbf{P}) = c)$ from the set $T2_k$ of multiple decision trees for which $G(k) = 1$ and use Eq. (4) to assign the keypoints of high probability into class c.

$$\hat{Y}(\mathbf{P}) = \arg\max_{c} \frac{1}{L_2}\sum_{l=1}^{L_2}\sum_{k=1}^{K} G(k)P_{\eta(T2_{kl},\mathbf{P})}(Y(\mathbf{P}) = c) \quad (4)$$

The proposed method classifies the input image viewpoints in the first stage RTs,and in uses the RTs that have been trained with images whose clasified viewpointin the second stage to achieve highly accurate corresponding point matching.

3 Evaluation Experiment

To show the effectiveness of the proposed method, we experimentally compared template corresponding point matching with the conventional method. We also conducted processing time experiments.

3.1 Database

We used the Mikolajczyk database[1] and the Morel database[2] in the experiments. The image data from the Morel database included seven images of an object rotated in the range from 10 degrees to 70 degrees (a) and four images of an object rotated in the range from 45 degrees to 80 degrees (b). The image data from the Mikolajczyk database included three images of an object rotated in the range from 10 degrees to 40 degrees (c).

3.2 Experiment overview

We compared SIFT[1], SURF[5] and Randomized Trees (RTs)[7] regarding the matching rate obtained from Eq. (5).

$$\text{Matching rate} = \frac{\text{Number of matching successes}}{\text{Number of matching}} \qquad (5)$$

The RTs were trained with 23,328 images, subsets for first stage and second stage RTs (L_1 and L_2) and a decision tree depth of 15.

3.3 Comparison with the conventional method

We conducted comparison experiments to test the effectiveness of the proposed method. We compared SIFT, SURF, RTs, and the proposed method for processing time. The personal computer used in the experiment had a Xeon@2.66 GHz processor.

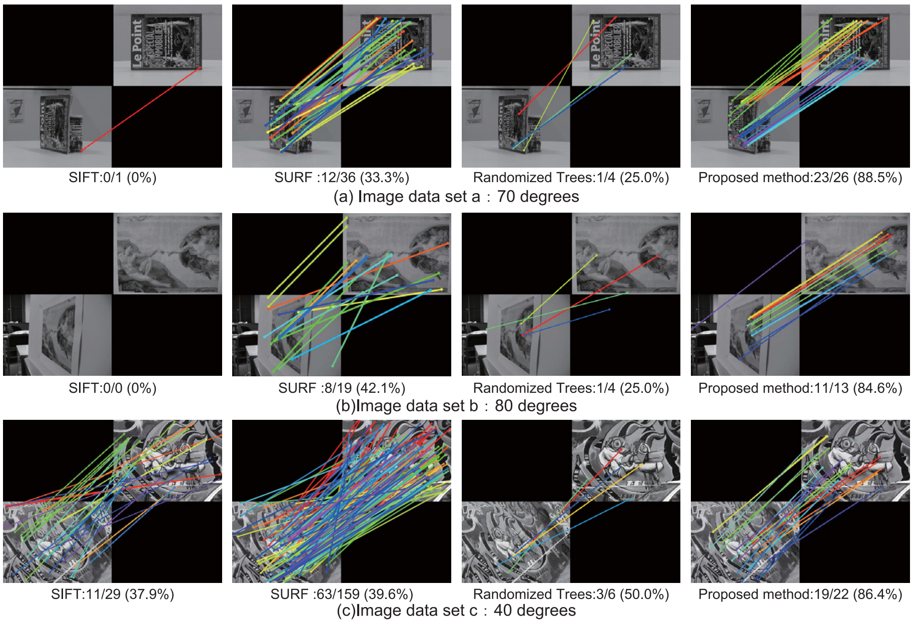
The matching accuracy results are presented in Table 1 through Table 3 for the various sets of image data, and The keypoint matching result image for each image data set is shown in Fig. 5. Moreover, the experimental results are presented in Fig. 4.

From the results presented in Tables 1 through 3, we know that the highest accuracy is achieved by the proposed method, followed by RTs, SURF, and finally SIFT. Compared to the conventional RTs method, the proposed method is more robust to changes in the image. The reason for this improvement is that, by training the RTs in stage 2, the many different viewpoints of the template can be limited in the first stage, simplifying the keypoint classification problem for the second stage RTs and thus improving accuracy. Moreover, false matching can be decreased using a homography matrix that is calculated using RANSAC. The RANSAC stably solves the problem using more number of the correct matching. It is possible to reduce the false matching using the homography matrix, because more number of the correct matching can be obtained by the proposed method.

The conventional RTs method has the fastest processing time, followed by the proposed method, SURF, and then SIFT. The proposed method uses two-stage RTs for matching, which increases the processing time by a factor of 1.7 relative to the conventional RTs method. Nevertheless, it is still capable of real-time processing at 12 fps.

[1] `http://www.robots.ox.ac.uk/~vgg/research/affine/`

[2] `http://www.cmap.polytechnique.fr/~yu/research/ASIFT/`



SIFT:0/1 (0%) SURF :12/36 (33.3%) Randomized Trees:1/4 (25.0%) Proposed method:23/26 (88.5%)

(a) Image data set a : 70 degrees

SIFT:0/0 (0%) SURF :8/19 (42.1%) Randomized Trees:1/4 (25.0%) Proposed method:11/13 (84.6%)

(b)Image data set b : 80 degrees

SIFT:11/29 (37.9%) SURF :63/159 (39.6%) Randomized Trees:3/6 (50.0%) Proposed method:19/22 (86.4%)

(c)Image data set c : 40 degrees

Figure 5. Keypoint matching results.

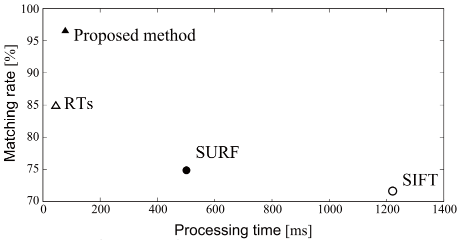


Figure 4. Relation between processing time and accuracy.

Table 1. Image data sets a matching rate [%].

	30	50	70	Avg.
SIFT	100.0	96.9	0.0	82.0
SURF	98.1	87.3	33.3	82.1
RTs	100.0	98.5	25.0	88.7
Proposed method	100.0	99.4	88.5	98.2

Table 2. Image data sets b matching rate [%].

	45	65	80	Avg.
SIFT	82.8	77.1	0.0	52.1
SURF	94.4	81.1	42.1	72.3
RTs	100.0	100.0	25.0	82.5
Proposed method	100.0	100.0	84.6	97.4

Table 3. Image data sets c matching rate [%].

	10	20	40	Avg.
SIFT	100.0	88.6	37.9	75.5
SURF	91.5	77.9	39.6	70.0
RTs	100.0	91.8	50.0	80.6
Proposed method	100.0	93.1	86.4	93.1

4 Conclusion

We have proposed here a keypoint classification method that uses two-stage Randomized Trees. That method represents the two problems of changes in template viewpoint and keypoint classification with two-stage Randomized Trees, which simplifies the classification problem compared to the conventional RTs method. The result is that even if the viewpoint of the target object is rotated by 70 degrees in the input image, an improvement in accuracy of 88.4% relative to SIFT and 63.4% relative to RTs is achieved. We confirmed that the proposed method is capable of corresponding point matching at 12 fps. In future work, we will investigate techniques for training RTs with less memory and on-line training methods.

References

[1] D. G. Lowe: "Distinctive image features from scale-invariant keypoints", Int.Journal of Computer Vision, **60**, pp. 91–110 (2004).

[2] Y. Ke and R. Sukthankar: "PCA-SIFT : A more distinctive representation for local image descriptors", Proc. of IEEE Conference on Computer Vision and Pattern Recognition, **2**, pp. 506–513 (2004).

[3] K. Mikolajczyk and C. Schmid: "A performance evaluation of local descriptors", IEEE Transactions on Pattern Analysis and Machine Intelligence, **27**, 10, pp. 35–47 (2005).

[4] S. Belongie, J. Malik and J. Puzicha: "Shape matching and object recognition using shape contexts", IEEE Transactions on Pattern Analysis and Machine Intelligence, **2**, 4, pp. 509–522 (2002).

[5] H. Bay, T. Tuytelaars and L. V. Gool: "SURF:speeded-up robust features", In ECCV, pp. 404–417 (2006).

[6] G. Michael, G. Helmut and B. Horst: "Fast approximated SIFT", Proc. of ACCV, pp. 918–927 (2006).

[7] V. Lepetit and P. Fua: "Keypoint recognition using randomized trees", Transactions on Pattern Analysis and Machine Intelligence, **28**, 9, pp. 1465–1479 (2006).

[8] L. Breiman: "Random forests", Machine Learning, 45(1), pp. 5–32 (2001).

[9] M. Ozuysal, M. Calonder, V. Lepetit and P. Fua: "Fast keypoint recognition using random ferns", IEEE Transactions on Pattern Analysis and Machine Intelligence (2009).

[10] D. Wagner, G. Reitmayr, A. Mulloni, T. Drummond and D. Schmalstieg: "Pose tracking from natural features on mobile phones", Proc. ISMAR 2008 (2008).