Information Theoretic Geometric Features Selection for 3-D Object Recognition

Shintarou Matsuzaki, Makoto Maeda and Katsuhiro Inoue
Department of Systems Design and Informatics, Kyushu Institute of Technology
680-4, Kawazu, Iizuka, Fukuoka 820-8502, Japan
E-mail: matsuzaki@kiri.ces.kyutech.ac.jp

Abstract
In this paper, we introduce a strategy which chooses significant features based on information theory in 3-D object recognition. Our optimality criterion is a reduction of uncertainty in a recognition process. If uncertainty and ambiguity of the recognition process can be reduced, object recognition becomes more reliable. A technique to choose the optimal feature based on information theory is already studied for active object recognition. This paper proposes a feature selection strategy for recognizing 3-D objects by extending such a framework. The strategy is constructed by an entropy-based approach using an iterative algorithm. Significant features is chosen based on a set of geometric features consisting of three features in the feature selection strategy. We present 3-D object recognition process, and discuss the validity of the proposed feature selection strategy via some experiments.

1 Introduction
3-D object recognition is one of the most attractive issues in computer vision. Especially the recognition techniques based on local features have been widely studied. In general, it is realized from 3-D data by extracting some geometric features such as feature points (key-points) and their local descriptors that describe the neighbor shapes around the key-points, so that both objects registered in a database and an observed scene can be represented by sets of these features. Examples are Scale-Invariant Feature Transform (SIFT) [6] of 2-D and Spin Image [4] of 3-D. Hence, the recognition of an object in a scene amounts to the discovery of a match between a subset of the scene features and a subset of the object features stored in the database. However, similar local descriptors are easily generated when lots of key-points are detected. In this case, since many matches to several objects are discovered, it tends to be difficult to decide an unique object from the matching votes.

Therefore the purpose of this study is to realize efficient recognition by features that provide more information quantities from viewpoint of information theory. The similar approaches have been developed to select an optimal camera position sequence for realizing most rapidly the recognition [1, 2]. This paper proposes the feature selection strategy for recognizing 3-D objects by extending such the framework.

The strategy is constructed by the entropy-based approach using the iterative algorithm. First, we choose an arbitrary ordered triplet of non-collinear points $A$, $B$ and $C$ from the key-points in the observed scene, and compute a set of geometric features from the triplet. By referring to the database, we can obtain some information on respective likelihoods to the registered objects. If all the likelihoods are not sufficiently high to identify the scene object, we select another triplet based on the information theory to obtain the set of significant features. This procedure is iterated until we can uniquely identify a right object with high certainty. Hence, the problem is to select the triplet to reduce most rapidly the uncertainty and the ambiguity of the recognition.

The paper is structured as follows: Section 2 describes local descriptors and the triplets. They are used for features selection and recognition in Section 3. Section 4 shows experimental results that applied the features selection strategy to 3-D object recognition. We summarize the experimental evaluation in Section 5, and then conclude with a discussion of results and a perspective on future work.

2 Features descriptor
In this section, we explain a method for extracting some geometric features such as the key-points and their local descriptors. First, the key-points are detected from the 3-D data of an object by using Harris operators [5]. We set the maximum number of the key-points to 50 for every object in this experiment. We choose an arbitrary ordered triplet of non-collinear points $A$, $B$ and $C$ from the key-points, so that a triangle $ABC$ is formed. We define the edge lengths
Fig. 1 shows relations of the triangle ABC. Arbitrary 3 points constituting the triplets are chosen among 50 key-points. Therefore there are 19,600 patterns to the triplets.

Next, the local descriptors are generated by using Point Feature Histograms (PFH) [3] in order to describe the neighbor shapes around the key-points. The PFH formulation is to encode a point’s k-neighborhood geometrical properties using a multi-dimensional histogram. Since k is set to 10, at most 10 points are searched from the neighborhood with distances smaller than a radius. The PFH is based on the relationships between the points in the k-neighborhood and their estimated surface normals. The surface normal is estimated at every 3-D data point of the object. The final PFH descriptor is computed as a histogram of relationships between all pairs of points in the neighborhood. To compute the relative difference between two points \( p_s \) and \( p_t \) and their associated normals \( \mathbf{n}_s \) and \( \mathbf{n}_t \), we define a fixed coordinate frame at one of the points.

By using a local coordinate frame, the difference between the two normals \( \mathbf{n}_s \) and \( \mathbf{n}_t \) can be expressed as a set of angular features as follows:

\[
\alpha = \mathbf{v} \cdot \mathbf{n}_t \\
\phi = \mathbf{u} \cdot \frac{(\mathbf{p}_t - \mathbf{p}_s)}{d} \\
\theta = \arctan(\mathbf{w} \cdot \mathbf{n}_t, \mathbf{u} \cdot \mathbf{n}_t)
\]

where \( d \) is the Euclidean distance between the two points \( \mathbf{p}_t \) and \( \mathbf{p}_s \). The quadruplet \( <\alpha, \phi, \theta, d> \) is computed for each pair of two points in k-neighborhood, therefore reducing the 12 values of the two points and their normals to 4. To create the final PFH representation for the query point, the set of all quadruplets is binned into a histogram. The PFH generates 125 dimensions of histogram. We use a description method of this PFH for the local descriptors.

However, we use Principal Component Analysis (PCA) compression because the number of the dimensions is high. Since we project the local descriptors on a low dimensional space.

Fig. 2 shows the relations of \( p_s \) and \( p_t \). Using a \(uvw\) coordinate, the difference between the two normals \( \mathbf{n}_s \) and \( \mathbf{n}_t \) can be expressed as a set of angular features as follows:

\[
u = \mathbf{n}_s \\
v = \mathbf{u} \times \frac{(\mathbf{p}_t - \mathbf{p}_s)}{\left\|\mathbf{p}_t - \mathbf{p}_s\right\|_2} \\
w = \mathbf{u} \times \mathbf{v}
\]
Fig. 4: An appropriate selection of the triplets $a_t$ in the iterative algorithm makes the estimated probability distribution $p(x_t)$ of the state $x_t$ have a unimodal distribution with small variance.

subspace by using PCA, each key-point has a compressed local descriptor. Hence, by joining the compressed local descriptors at every vertex of the triplet, the triplet can be also described as a joint feature $o$.

As shown in Fig.3, each object is registered in the database as a set of triplets that are also described by the geometric relations $a$ and the joint features $o$. Thus the object recognition is realized by discovering correspondences among the triplets.

3 Recognition process

In this section, we explain the recognition process using the iterative algorithm. The object recognition can be formulated as a state estimation problem, in which the state $x_t = [x^1_t, x^2_t, \ldots, x^N_t]$ ($N$ denotes the number of objects) stands for the class of an object estimated at time step $t$ in the iterative algorithm. In our method, the state estimation is performed by an entropy-based approach using an iterative algorithm. The state to be estimated is described with a probability density function $p(x_t)$ at time step $t$. The pdf indicates a certainty of the state. At the first step in the iterative algorithm, since we have no knowledge about an object, the pdf is considered as a uniform distribution as shown in the left-hand side of Fig.4. However, as the algorithm is iterated, the pdf may be a unimodal distribution with small variance as shown in the right-hand side of Fig.4. Therefore, we can uniquely identify the right object.

In the iterative algorithm, we try to choose the triplets one after another so as to transform the pdf $p(x_t)$ to a unimodal distribution with small variance. the joint feature $o_t = [o^1_t, o^2_t, \ldots, o^n_t]$ ($n$ denotes a dimension compressed in PCA) is provided by choosing the geometric relation $a_t$. We calculate the joint feature from chosen the triplets, and recognize the object using the probability recognition model. We learn it beforehand to perform the feature selection and the object registration. A training stage is performed as follows.

First, we calculate a model of a noise component in the observation of the joint feature $o_t$. At time step $t$ when the joint feature $o_t$ is obtained with a geometric relation $a_t$ by selecting one of triplets that are generated from an object $x_t$, we can determine the likelihood function $p(o_t|x_t, a_t)$ as shown in Fig.5. Moreover, we can also determine the likelihood function $p(o_t|a_t)$.

Next, we calculate a geometric relations model that is computed from the triplets of an object $x_t$. We determine the likelihood function $p(a_t|x_t)$ from distribution of a geometric relation $a_t$ by selecting one of triplets. Moreover, we can also determine the likelihood function $p(a_t)$. The training stage is finished in the above. A recognition stage is performed as follows.

Fig. 5: Evaluation of likelihood values of the joint features for every objects.

- 25 -
Fig. 6: Change of \( p(x_t) \) by selecting the triplet \( a_t \) and obtaining the joint feature \( o_t \) at \( t \)-th step.

We explain in detail the selection method of triplets based on information, and describe the probabilistic recognition model for the object recognition.

### 3.1 Features selection based on information theory

The efficient triplets for the recognition are chosen based on information theory. Our theoretical base is that the uncertainty of the recognition can be efficiently minimized by selecting the triplets. We consider an entropy \( H(x_t) \) that is associated with the pdf \( p(x_t) \). The entropy measures the amount of uncertainty in \( p(x_t) \).

The entropy \( H(x_t) \) is defined as

\[
H(x_t) = - \sum_{i=1}^{N} p(x_t^i) \log p(x_t^i). \tag{2}
\]

It is zero if the outcome of the recognition is unambiguous; it reaches its maximum if all outcomes of the recognition are equally likely. In information theory, conditional mutual information \( I(x_t; o_t | a_t) \) defines how much the uncertainty is reduced in \( x_t \) if the joint feature \( o_t \) is obtained with the geometric relation \( a_t \).

The conditional mutual information is defined as

\[
I(x_t; o_t | a_t) = H(x_t) - H(x_t | o_t, a_t) = H(o_t) - H(o_t | x_t, a_t). \tag{3}
\]

Using an above notation for conditional probabilities and the definition of the entropies \( H(x_t) \) and \( H(x_t | o_t, a_t) \), the conditional mutual information becomes

\[
I(x_t; o_t | a_t) = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{n} p(x_t^i) p(o_t^j | x_t^i, a_t) \log \frac{p(o_t^j | x_t^i, a_t)}{p(o_t^j | a_t)}. \tag{4}
\]

Since we are interested in reducing the uncertainty, we have to maximize the mutual information \( I(x_t; o_t | a_t) \). Since the mutual information is a function of the parameter \( a_t \), the optimal parameter \( a_t^* \) is found as

\[
a_t^* = \arg \max_{a_t} I(x_t; o_t | a_t). \tag{5}
\]

Once the optimal parameter \( a_t^* \) is found, that allows us to select one from the triplets that have the geometric relation \( a_t^* \). This means that, by employing the mutual entropy \( I(x_t; o_t | a_t) \) as the criterion to select the geometric relation \( a_t \), we can evaluate the likelihood of the observed object.

### 3.2 Probabilistic recognition model

This section explains a probabilistic recognition model. we can evaluate a posteriori probability \( p(x_t | o_t, a_t) \) since it is modified by Bayes’ theorem as follows.

![Data set of 3-D objects for classification.](image-url)
Table 1: Results for the proposed features selection and random selection(100 Trials for each object).

<table>
<thead>
<tr>
<th>object</th>
<th>features selection</th>
<th>random selection</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>recognition rate(%)</td>
<td>mean steps</td>
</tr>
<tr>
<td>$x^1$</td>
<td>88</td>
<td>18.8</td>
</tr>
<tr>
<td>$x^2$</td>
<td>84</td>
<td>16.9</td>
</tr>
<tr>
<td>$x^3$</td>
<td>98</td>
<td>4.3</td>
</tr>
<tr>
<td>$x^4$</td>
<td>65</td>
<td>14.3</td>
</tr>
<tr>
<td>$x^5$</td>
<td>92</td>
<td>9.1</td>
</tr>
<tr>
<td>average</td>
<td>85.4</td>
<td>12.6</td>
</tr>
</tbody>
</table>

\[
p(x_t|\alpha_t, a_t) = \frac{p(\alpha_t|x_t, a_t)p(x_t|a_t)}{p(\alpha_t|a_t)}
\]

\[
= \frac{p(\alpha_t|x_t, a_t)p(a_t|x_t)p(x_t)}{p(\alpha_t|a_t)p(a_t)}.
\] (6)

Fig.6 shows update of $p(x_t)$. The object recognition is just to evaluate $p(x_t|\alpha_t, a_t)$ for every object. The posteriori probabilities can be interpreted as new a priori probabilities for the next step, i.e., $p(x_t) = p(x_{t-1}|\alpha_{t-1}, a_{t-1})$.

Consequently, we can realize the object recognition with selecting the significant features.

### 4 Experimental results

In order to verify the feasibility of our strategy, we performed the recognition experiments using 5 objects. The objects to be recognized are shown in Fig.7.

At first we calculate the likelihood functions for the object recognition in the training stage. We detect 50 key-points for each object. Therefore the geometric relations $a_t$ that have 19,600 patterns exist. The most important part in this stage is to estimate the conditional density $p(\alpha_t|x_t, a_t)$. $p(\alpha_t|x_t, a_t)$ is modeled using the joint features generated among the key-point and its neighbor 5 points for each object. Other likelihood functions are calculated as mentioned in Sec.3. These procedures are performed in off-line. When the training stage is finished, we move to the recognition stage.

In the recognition stage, we find 50 key-points from an observed object. A triplet is randomly selected at the initial step. We calculate $p(x_t|\alpha_t, a_t)$ for each object from the triplet and the joint future. We decided the target object to be object $x^T$ if the likelihood or the normalized probability becomes larger than their respective thresholds, that is:

$p(x^T_t|\alpha_t, a_t) > 0.4$, and $\frac{p(x^T_t|\alpha_t, a_t)}{\sum_{x} p(x_t|\alpha_t, a_t)} > 0.9$.

Otherwise, we select next other triplet according to the proposed strategy, and continue the process. Fig.8 shows the flowchart of the object recognition process. The experiments were performed as follows:

1. **Initialization**
   
   Initialize the pdf $p(x_t)$ uniformly.

2. **Features selection**

   Find a parameter which maximizes the mutual information criterion (5). The parameter indicates the most effective geometric relation.

3. **Feature extraction**

   Choose one from the triplets that have the most effective geometric relation. Calculate the joint feature based on the triplet.

4. **Bayes decision**

   Calculate $p(x_t|\alpha_t, a_t)$ using Bayes formula for each object.

5. **Loop or end**

   If the a posteriori probability for one class satisfied with the termination condition, then end. Otherwise, set the a priori probability for the next time step to the current a posteriori probability. Go to 2.

To compare with the proposed strategy, we also carried out the object recognition process using a random selecting...
approach. In 2. Features selection and 3. Feature extraction of Fig.8, triplets were randomly selected.

The recognition experiments were performed 100 times using the two approaches: the proposed features selection approach and the random selection approach. Table 1 shows the recognition rate and the number of steps to reach the final decision, where these values are the mean values of 100 trials. From this table, we can confirm that the recognition using the proposed features selection has higher recognition rate than the random selection for all objects. This result indicates the effectiveness of the proposed strategy.

![Rate vs. Number of Steps](image.png)

**Fig. 9:** The number of steps required to reach the final decision by the proposed strategy and the random selection of the triplet.

We also discuss the number of steps. Fig.9 shows how many time was required to reach the final decision by the proposed features selection and the random selection. Although Fig.9 shows the change of the probability in recognizing the object $x^1$, the similar results were also obtained in recognizing other objects. We confirmed the proposed strategy can reduce the number of steps on the average to reach the final decision. However, the random selection requires less mean steps than the proposed features selection for object $x^2$, $x^3$ and $x^5$. In some experiments, there were the case that the proposed strategy required the large steps until reaching the final decision. In most of such cases, the proposed process failed the recognition. Although it calls for further investigation, it is clear that it leads to increase the mean steps.

Next, we discuss the cost of computation. The computation time to reach final decision depends on the number of triplets $a$, the number of joint features and the number of objects (classes). In the case of the differential mutual information, the computation time depends on the number of steps, the number of objects, and the number of samples taken to approximate the differential mutual information. In future work, we will investigate the predominancy of such a total cost.

**5 Conclusion**

State estimation is a formalism that can be used to frame the most important problems in computer vision. Clearly, the observations (images, features, high-level structures) have a strong influence on the accuracy of state estimation. Thus, either implicitly or explicitly most systems cycle through a state estimation and selection stage.

We have proposed a feature selection strategy for recognizing 3-D objects, in which it is constructed by an entropy-based approach using an iterative algorithm. The feasibility of the proposed strategy was verified by experiments. That enables us to realize the effective recognition of objects.

**Acknowledgement**

This work has been supported by JSPS KAKENHI Grant Number 24500209.

**References**


