Incremental Learning of Visual Landmarks for Mobile Robotics

Antonio Bandera, Rebeca Marfil
ISIS, Dpto. Tecnología Electrónica, Universidad de Málaga
{ajbandera, rebeca}@uma.es

Ricardo Vázquez-Martín
CITIC, Parque Tecnológico de Andalucía
rvazquez@citic.es

Abstract

This paper proposes an incremental scheme for visual landmark learning and recognition. The feature selection stage characterises the landmark using the OpponentSIFT, a color-based variant of the SIFT descriptor. To reduce the dimensionality of this descriptor, an incremental non-parametric discriminant analysis is conducted to seek directions for efficient discrimination (incremental eigenspace learning). On the other hand, the classification stage uses the incremental evolving clustering method (ECM) to group feature vectors into a set of clusters (incremental prototype learning). Then, the final classification is conducted based on the k-nearest neighbor approach, whose prototypes were updated by the ECM. This global scheme enables a classifier to learn incrementally, on-line, and in one-pass. Besides, the ECM allows to reduce the memory and computation expenses. Experimental results show that the proposed recognition system is well suited to be used by an autonomous mobile robot.

1. Introduction

In the mobile robotics community, there exists a recent interest for using object recognition algorithms to provide natural landmarks for the sake of simultaneous robot localization and environment mapping. In this framework, visual landmarks are detected little by little and the properties of the scenario where they are acquired could be slightly changed as time passes. Therefore, the landmark recognition system learning process should be also conducted sequentially in an on-line manner. Besides, it is desirable that the human supervisor only provides training samples to the robot when it does not correctly classify autonomously perceived patterns. Incremental learning is primarily focused on processing the data in a sequential way so that in the end the classifier is no worse than a hypothetical classifier trained on the batch data [1].

This paper describes an incremental scheme for visual landmark recognition which can perform without an a priori knowledge about the number and shape of the classes which compound the feature space. Following previous related work [1], this incremental approach combines two learning schemes which are, in this case, an incremental non-parametric discriminant analysis (INDA) [5] and the evolving clustering method (ECM) [3]. Fig. 1 shows an overview of the proposed system. Briefly, the on-line process of classification and learning is conducted as follows. First, N training samples must be given in advance to form an initial eigenspace model. The ECM is used to cluster the transformed versions of these samples into a set of groups represented by their prototypes. When a query input is presented to the system, the classification is carried out. The input descriptor is projected into the current eigenspace to obtain a reduced feature vector. The classification of this vector is conducted by a k-nearest neighbor algorithm whose prototypes are provided by the ECM. If the training sample is misclassified, this input and its class label are applied to the INDA to update the current eigenspace model. The updated eigenspace model is used for transforming the query input into a new feature vector, and this vector as well as the updated prototypes are used to train the classifier using the ECM. Thus, the proposed approach is able to simultaneously perform the feature selection and the classifier learning in one-pass, being training samples presented only once to learn [1].

This paper is organized as follows. Section 2 presents the two stages of the proposed method. The experimental results revealing the performance of the
method are described in Section 3. The paper concludes along with discussions and future work in Section 4.

2 Proposed approach

2.1 Feature selection stage

Visual landmarks can be defined as distinct environment features that can be recognized reliably from sensor observations. Hence, they could be associated to surface patches which are significantly different from its surroundings (salient image regions). In this proposal, landmarks are obtained using a novel approach for affine, salient region detection which looks for high-contrasted regions of data-dependent shape. Fig. 2 shows the sets of visual landmarks extracted from several images which represent the same scene observed from different viewpoints. It can be noted that these sets are very similar. More extensive tests have shown that this detector reliably finds the same visual landmarks under different viewing and illumination conditions [7].

On the other hand, image regions are described in this approach using a color-based variant of the SIFT descriptor [4]. If the SIFT descriptor encodes the image region employing edge orientation histograms computed over the intensity image, the color-based variants compute these histograms over the different channels used to encode the color information. Among the different color SIFT descriptors, we have employed in our approach the OpponentSIFT [6]. This descriptor encodes all the channels of the opponent color space. This color space is defined as

$$\begin{pmatrix} O_1 \\ O_2 \\ O_3 \end{pmatrix} = \begin{pmatrix} \frac{R-G}{G+B} \\ \frac{R+G-2B}{G+B} \\ \frac{2\sqrt{2} R + B}{\sqrt{3}} \end{pmatrix}$$

(1)

being the color information represented in channels $O_1$ and $O_2$, and the intensity in channel $O_3$. As SIFT descriptors are computed using 128 dimensions, the OpponentSIFT descriptor has $3 \times 128$ dimensions. To reduce its size, descriptors are transformed to a more reduced feature space using an Incremental Non-parametric Discriminant Analysis (INDA) approach.

Let us assume that $N$ training descriptors $\{x_i\}_i$ belonging to $M$ classes $C_i$ have been presented. Let us assume that each class $C_i$ is composed by $m_i$ samples, $C_i = \{x_1^i, x_2^i, ... x_{m_i}^i\}$, where $x_i^i$ is the mean vector of class $C_i$. Non-parametric discriminant analysis (NDA) seeks a transformation $U$ over the set of training samples in such a way that the ratio of the between-class matrix $S_b$ and the within-class matrix $S_w$ is maximized. In the non-parametric scheme, these matrices are defined as [2]

$$S_b = \sum_{i=1}^{M} \sum_{j=1}^{M} m_i \sum_{t \neq i} w_{C_i, C_j}^T (x_t^i - \mu_{C_i}(x_t^i))(x_t^i - \mu_{C_j}(x_t^i))^T$$

(2)

$$S_w = \sum_{i=1}^{M} \sum_{j \in C_i} (x_j^i - \bar{x}^i)(x_j^i - \bar{x}^i)^T$$

(3)

where $\mu_{C_j}(x_t^i)$ is the average of the $k$-nearest neighbor of $(x_t^i)$ in class $C_j$. $w_{C_i, C_j}^T$ is a weighting function which de-emphasizes the contribution of samples that lie far from the class boundary [2].

Assuming that these matrices have been computed from at least two classes, let us consider that a new training sample $y$ is presented to the system. Two situations can be distinguished [5]:

- If the input sample $y$ belongs to the class $C_L$, then the between-class scatter matrix will be updated by removing the scatter matrix $S_b^{in}(C_L)$ associated to the old class $C_L$, and adding a new scatter matrix associated to the class $C_{L'} = C_L \cup y$. Then, the
Figure 2. Visual landmarks extracted from several images which represent the same scene from different viewpoints. Representing ellipses have been chosen to have the same first and second moments as the originally arbitrarily shaped regions.

new between-class scatter matrix is defined by

$$S'_b = S_b - S_b^{in}(C_L) + S_b^{in}(C_{L'}) + S_b^{out}(y)$$ (4)

where the term $S_b^{out}(y)$ is defined by

$$S_b^{out}(y) = \sum_{i=1, i \neq L}^M (y - \mu_{C_i(y)})(y - \mu_{C_i(y)})^T$$ (5)

On the other hand, to estimate the new within-class scatter matrix, we only need to calculate the scatter matrix of the new class, $S_w(C_{L'}):

$$S_w(C_{L'}) = S_w(C_L) + \frac{m_L}{m_L + 1}(y - \bar{x_L})(y - \bar{x_L})^T,$$ (6)

and replace the old value $S_w(C_L)$ in the estimation of the new $S'_w$.

- If the input sample $y$ belongs to a new class $C_{M+1}$, the between-class scatter matrix is updated using the following equation

$$S'_b = S_b + S_b^{out}(y) + S_b^{in}(C_{M+1})$$ (7)

In this case, the within-class scatter matrix remains unchanged ($S'_w = S_w$).

Once the scatter matrices are obtained, the transformation matrix $U$ can be computed by conducting an eigenvalue decomposition of the matrix $D = S_w^{-1}S_b$,

$$DU = UA$$ (8)

The columns of the transformation matrix correspond to the discriminant eigenvectors. In our tests, a reduced set of eigenvectors is chosen to ensure that the projection of the data onto this set covers at least 90% of the data’s spread. It must be noted that the number of eigenvectors to maintain a constant value of coverage of the data’s spread could change during the incremental eigenspace learning.

2.2 Classification stage

The classification stage combines the ECM with a $k$-nearest neighbor algorithm. The ECM is a fast one-pass, distance-based clustering algorithm. It is able to include the input feature vector in the most suitable existing cluster, creating a new one if necessary, with no supervision nor previous training [3]. In our on-line classifier, this algorithm is used to cluster the input patterns into a reduced set of groups, which are defined by their cluster centers ($prototypes$). That is, in our approach, the ECM does not need to store all input samples but only the prototype of every group ($prototypes$). Thus, it reduces the memory and computational expenses of the whole approach. On the contrary, all input samples cannot be used to redefine the set of data groups (reference set) when a new sample arrives.

The prototypes are determined such that the maximum distance between an input sample and the closest prototype cannot be larger than a threshold value, $T$. In our tests, this threshold will be set to a low value, as it is only employed to reduce the number of stored items, but not to provide the final classification. Once the set of prototypes has been updated, the $k$-nearest neighbor is employed to classify the input sample. Euclidean distance can be used to measure the similarity between the transformed versions of the input descriptor $y$ and each stored prototype $\bar{x}_j$, which is also an OpponentSIFT descriptor. As in the ECM, this distance $d_{y\bar{x}_j}$ is computed as

$$d_{y\bar{x}_j} = ||U^Ty - U^T\bar{x}_j||_2,$$ (9)

where $U$ is the transformation matrix provided by the INDA. It can be noted that stored prototypes are OpponentSIFT descriptors. When the transformation matrix $U$ is improved because more samples have been processed, classification results are also better.
3 Experimental results

The experimental evaluation has been conducted on a Pioneer 2AT platform from ActivMedia. The image acquisition system used in the experiments employs a STH-MDCS stereoscopic camera from Videre Design. Images were restricted to $320 \times 240$ pixels.

Fig. 3 shows several visual landmarks obtained from a typical trial. In this experiment, a first chunk of 200 samples were employed to build the initial feature eigenspace through conventional NDA. In this initial chunk only 16 different visual landmarks were included. Besides, a set of 40 samples was used to define a 'background' group. Then, 528 new images were acquired and processed. From these images, 2670 tentative landmarks were extracted and characterized. They were manually labelled in order to estimate the recognition accuracy. However, the robot does not have this information. On the contrary, it must determine the confidence of each recently classified landmark. Thresholding this value, it is able to classify a landmark as an 'unrecognized' one, requiring the help of the user. The confidence value $C$ is defined as

$$
C = \min \{ C_i \} \cap_{i=0}^{n_p} = \min \left( 1 - n_p \frac{d(y \bar{x}_i)}{\sum_{j=0}^{n_p} d(y \bar{x}_j)} \right) \cap_{i=0}^{n_p},
$$

(10)

where $C_i$ displays the similarity of an input sample $y$ with the $i$ prototype stored in the system, $\bar{x}_i$. An input sample $y$ is considered to be classified as 'unrecognized' if the obtained confidence value $C$ is under a fixed threshold, which has been empirically set to 0.25 in our tests. From the total set of detected landmarks, a set of 1237 landmarks was classified on 34 different groups. The rest of extracted landmarks was classified as 'background'. Fig. 4 shows the number of landmarks which were supervisely classified (red line). It can be noted that the detection of 'unrecognized' landmarks decreased as the robot travel through the office-like environment, where the same landmarks (doors, lights, windows...) are repeatedly found. Fig. 4 also shows the evolution of the recognition accuracy (blue line). Using a Pentium IV 2.6 GHz PC, the process of detection and classification takes less than 0.5 second for each $320 \times 240$ image.

4 Conclusions

This paper has described a visual landmark recognition system. The representation stage employs the OpponentSIFT algorithm to describe each visual landmark. To simultaneously conduct feature selection and classifier learning, the approach uses a global incremental scheme which combines the INDA and the ECM approaches. This scheme enables a classifier to learn incrementally, on-line, and in one-pass. Experimental results show that the system works correctly with no a priori information about the nature of the input patterns. Future works will be focused on improving the current implementation in order to increase its speed.

References