Interactive Machine Learning tool with Automatic Tagging for Video Recognition System

Fujio Tsutsumi
Central Research Institute of Electric
Power Industry
2-11-1, Iwado Kita, Komae-shi,
Tokyo 201-8511 JAPAN
+81-3-3480-2111
tutumi@criepi.denken.or.jp

ABSTRACT
Interactive machine learning (IML) is proposed to train intelligent systems. This paper proposes a new approach to solve the feedback delay and low visibility problem of conventional IML handling large-scale data. The approach features high-speed feedback using an approximate nearest neighbor, process controllability by the user, and priority based visualization. The validity of the approach is shown by an experiment with a prototype using a video where the lighting changes.

ACM Classification: H5.2 [Information interfaces and presentation]: User Interfaces. - Graphical user interfaces.

General terms: Algorithms, Design, Human Factors

Keywords: Visual Data Mining, Information Visualization, Interactive Machine Learning, Video Image Processing

INTRODUCTION
Supervised machine learning is a powerful technique for building intelligent systems. It allows automatic creation of classifiers, however, the classical approach is generally slow to train, and not interactive. Therefore, preparation of appropriate training data takes time and effort. In order to reduce the cost, interactive machine learning (IML) has been proposed[2]. IML prepares the software environment where the user can label data easily (indicating training samples and instructing correct answers). Moreover, if the user changes the labels, the recognition result based on the labels is immediately visualized. The user can detect suitable training data which should be labeled by the direct manipulation style user interface. Although IML is a promising approach, if the data becomes large like multimedia data, quick feedback becomes difficult and the visibility of recognition result deteriorates. This paper proposes a new approach which has the following three features.

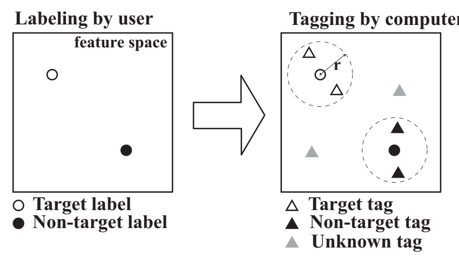

Figure 2: Concept of automatic tagging

- Feedback speed-up using LSH
- User controllable batch operation
- Visualization based on priority

PROTOTYPE OF INTERACTIVE VIDEO TAGGING TOOL
A prototype software for video data based on our approach describing at the following has been developed. Using the user interface, the user divides the frame into square blocks like a checkerboard. Each block is an unit for labeling and recognition. The prototype loads and plays a video file, and performs recognition with visualization. Figure 1 shows the visualization result of a 2 minute video.

The user can indicate the labels in a current displayed frame by mouse clicking targets. After the labeling, the computer guesses the label of data unlabeled. We call the guessed label a "**tag**". Tags have the following three states, "Target", "Non-target" and "Unknown". Target tags are attached to the data similar to true labeled data (see fig. 2). The tags contain many errors in the initial stage of learning. In IML, the user promotes learning of system by correcting the tags interactively.

When the user pushes the registration button, target labels and target tags are registered as target labels. On the other hand, non-target labels, non-target tags and unknown tags in the frame are registered as non-target labels automatically. So the user can register many labels by only few mouse clicks.

Training state of the system is visualized on real-time simultaneously with video play, and the user can pause and skip

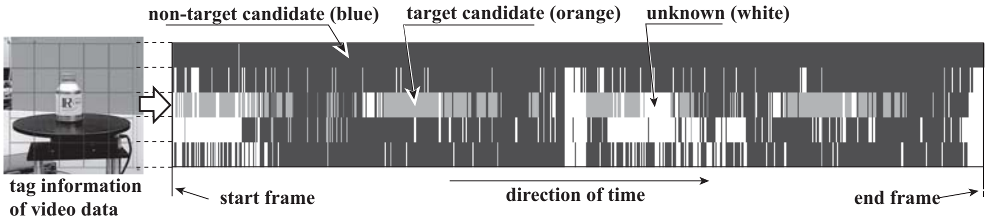

Figure 1: Recognition results of a frame are displayed as a vertical line of visualized time-line.

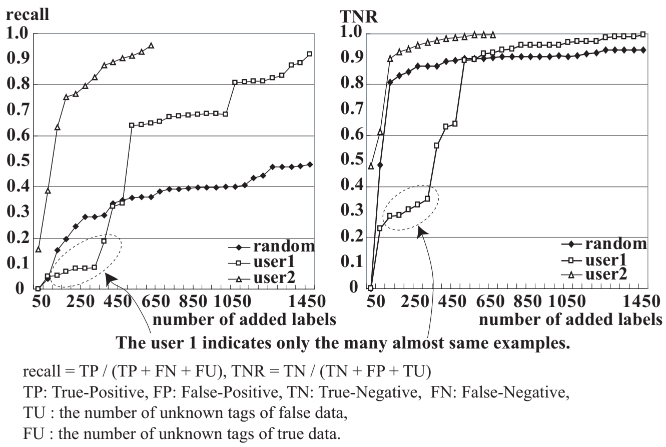

Figure 3: Experimental results of mark recognition.

the frame always.

EXPERIMENTAL RESULT

The prototype was evaluated using three sample videos (47sec, 2min, 1hour). In this paper, only the result of the shortest video which imitates lighting change (1424frames) is described. In the video the logo mark of our laboratory (with green, orange and blue parts) printed is photoed by 120×60 pixels. Lighting changes from fluorescent light, sunlight to shadow in this order. According to the lighting changes, the AE and AWB functions of the video camera work, and the appearance of the logo mark changes. The task is correct logo recognition under the lighting change. The used image feature is 8bit RGB bi-linear scaling data in 27 dimensions.

Two human subjects tried to teach the prototype. Figure 3 shows the result of experiments. For comparison with the subjects, automatic processing which carries out correct label attachment for random selected frames was performed ("random" in fig. 3). Since both FP and FN are extremely small in all of the results, precision $\cong 1$ and FPR $\cong 0$.

This result shows the effectiveness of the prototype, because both users have achieved high recall and TNR. However, there is a difference between the two results, and one of subjects showed lower recall rate than random selection in the initial stage, which means that efficiency depends on the user's focus and operational skill.

INTERACTIVE TAGGING TOOL USING LSH

High-speed feedback using LSH

Conventional IMLs adopt simple decision tree (DT) to learn the user's labels. However, the tree becomes imbalanced and, as a result the speed falls, when many labels are added continuously. This paper proposes using Locality-Sensitive Hashing (LSH)[1] for the learning mechanism, because LSH has the feature that registration speed and search speed do not fall for large-scale data. Experiments show that the LSH is up to 40 times faster than the kd-tree which is a typical DT[1]. Since our prototype can treat the general Euclid distance by image similarity, p-stable distribution type LSH has been used.

User controllable tagging

In the case of large amount of data, depending only on direct manipulation is impossible, because tag estimation takes much computational time. To cope with this problem, this paper proposes a user controllable automatic tagging according to a guideline as shown below.

- The user can interrupt tagging procedure and change the range of tagging at any time.
- Previously guessed tags are stored in the DB and displayed immediately according to the user's demand.

Tag visualization priority

Visualization of tags is an important element of IML, however, since there is a limit to the screen and vision, all of the tags can not be displayed. This paper introduces tag priority to solve this peep-hole problem.

Ordering of tag priority is Unknown>Target>Non-target, and tags are visualized according to this order. That is, when the same area should display two tags, the tag with the higher priority is displayed. By assigning the maximum priority to the "unknown" tag, the effective range of the label becomes clear. Unlike conventional IMLs, the unknown tag is utilized, since the labels are only available to part of the data.

REFERENCES

1. Mayur Datar, Nicole Immorlica, Piotr Indyk, and Vahab S. Mirrokni. Locality-sensitive hashing scheme based on p-stable distributions. In *Proceedings of the twentieth annual symposium on Computational geometry*, pages 253–262, 2004.

2. Jerry Alan Fails and Dan R. Olsen Jr. Interactive machine learning. In *Proceedings of the 8th international conference on Intelligent user interfaces*, pages 39–45, 2003.