A Fast Search Algorithm For Large Video Database Using HOG Based Features

Qiu Chen[1], Koji Kotani[2], Feifei Lee[3] and Tadahiro Ohmi[4]

[1]Department of information and Communication Engineering,
Kogakuin University, Japan
[2]Department of Electronics, Graduate School of Engineering,
Tohoku University, Japan
[3]School of Optical-Electrical and Computer Engineering,
University of Shanghai for Science and Technology, China
[4]New Industry Creation Hatchery Center, Tohoku University, Japan
`chen@cc.kogakuin.ac.jp, feifeilee1701@163.com`

ABSTRACT

In this paper, we propose a novel fast video search algorithm for large video database. Histogram of Oriented Gradients (HOG) has been reported which can be reliably applied to object detection, especially pedestrian detection. We use HOG based features as a feature vector of a frame image in this study. Combined with active search, a temporal pruning algorithm, fast and robust video search can be achieved. The proposed search algorithm has been evaluated by 6 hours of video to search for given 200 video clips which each length is 15 seconds. Experimental results show the proposed algorithm can detect the similar video clip more accurately and robust against Gaussian noise than conventional fast video search algorithm.

KEYWORDS

Fast search, Video database, HOG Features

1. INTRODUCTION

In recent years, video content becomes commonplace on the web and the size of video database quickly increases due to rapid developments of internet connection and disk storage technology in recent years. Video retrieval has become a hot area of research in current. Video search is an important problem in this area because it has a wide range of applications such as TV commercials detection [1], video copyright enforcement [2],[3], video clustering and so on. In this paper, video search means when a user presents a query video clip to the search engine, the search engine should identify all similar ones, that is to say, accurately locate the position of query video clip if it exists in the video database.

Many video search algorithms [7]-[10] have been proposed, and achieves successes to a certain extent. But such algorithms, however, are computational-power hungry for the exhaustive search of large video database. For large video database, Search speed is an important issue of video search. Base on active search [4], a temporal pruning algorithm, Kashino et al. [1] improved the conventional multimedia search algorithm. Nevertheless, their feature extraction utilizes intensity features of the frame image, so the results may be sensitive to small change of luminance and motion in the frame. In this paper, we utilizes a new feature based on Histogram of Oriented Gradients (HOG) [5] features, which had been reliably applied to object detection, especially pedestrian detection [6][11]. It has the following advantages: computational simplicity, motion-insensitivity and luminance-insensitivity. Because such a feature is compatible with active search algorithm, fast search speed can also be achieved by combining HOG based features and active search.
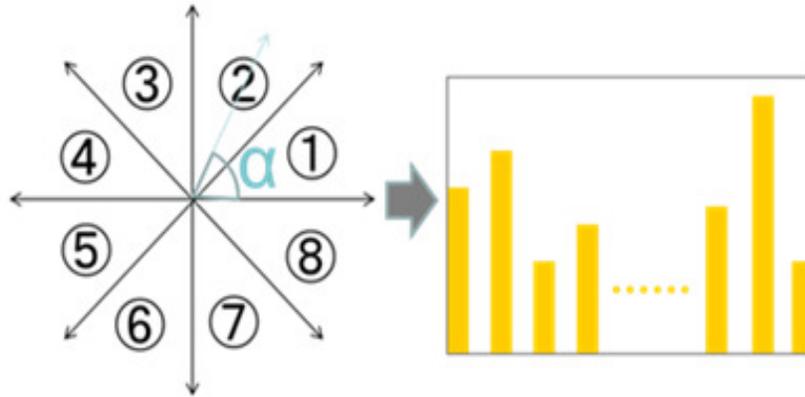
Figure 1. Processing steps of HOG based features.

In section 2, we will first introduce HOG based features, and then describe fast video search algorithm we employ in section 3. Experimental results compared to conventional search approach will be discussed in section 4. Finally, conclusions are given in section 5.

2. HOG BASED FEATURES

Histogram of Oriented Gradients (HOG) [5] features has been developed for object recognition previously. Figure 1 shows the processing steps of HOG based features. In HOG, for each pixel of an input image, the intensity difference of the horizontally adjacent pixels (fx) and the intensity difference of the vertically adjacent pixels (fy) are first calculated by using simple subtraction operations shown as formula (1), (2).

$$fy = f(i, j + 1) - f(i, j - 1) \tag{1}$$

$$fx = f(i + 1, j) - f(i - 1, j) \tag{2}$$

$$\alpha = tan^{-1} \frac{fy}{fx} \tag{3}$$

Each intensity variation vector is then quantized simply by the oriented angle calculated by formula (3). The number of vectors quantized in each quantization region is counted and a histogram is generated.

The essence of HOG based features above can be considered that the operation detects and quantizes the direction of intensity variation in the image blocks. Hence HOG based features contain very effective image feature information. We will describe how to apply it as feature vector of frame to solving the fast video search problem in next section.

3. PROPOSED VIDEO SEARCH ALGORITHM

The procedure of proposed fast search algorithm is shown in figure 2. In the pre-processing stage, the feature vectors are calculated from the query video clip and the video database HOG method described in section 2. The feature vectors are then quantized using VQ algorithm which applied quantization by combinations of scalar quantization (SVQ) for each feature dimension [1]. In the search stage, the windows are applied to both the query feature vectors and the feature vectors of video database. In the next step, the number of vectors quantized in the windows of the query video clip and video database are counted and feature vector histograms are created respectively. The similarity between these histograms is then calculated. If the similarity exceeded a threshold value given previously, the query video clip will be detected and located. Otherwise, the window on the video database will be skipped to the next position determined by the similarity in current position and the threshold value. In the last step, the window on the video database is shifted forward in time and the search proceeds.
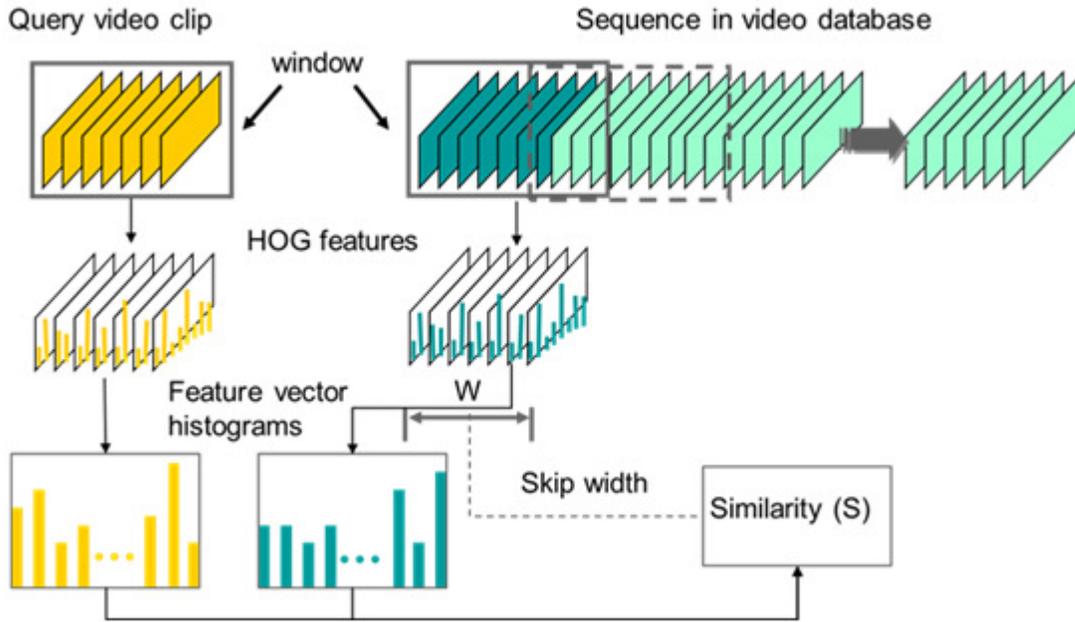
Figure 2. The procedure of proposed fast search algorithm

Here, histogram intersection is used as the similarity measure [4], and is defined as formula (4).

$$S_{QD} = S(H_Q, H_D)$$
$$= \frac{1}{N} \sum_{l=1}^{L} \min(h_{Ql}, h_{Dl}) \qquad (4)$$

where h_{Ql}, h_{Dl} are the numbers of feature vectors contained in the *l-th* bin of the histograms for the query and the stored signal, respectively, *L* is the number of histogram bins, and *N* is the total number of feature vectors contained in the histogram. The skip width *w* is shown by formula (5).

$$w = \begin{cases} floor(N(\theta - S_{QD})) + 1 & (S_{QD} < \theta) \\ 1 & otherwise \end{cases} \qquad (5)$$

where *floor(x)* means the greatest integral value less than *x*, and θ is a given threshold.

4. EXPERIMENTAL RESULTS AND DISCUSSIONS

We performed all of the experiments on a conventional PC @ 3.2GHz (4G memory). The algorithm was implemented in ANSI C. We used 6 hours of video captured from TV program. In the experiment, the video frame rate was 14.97 fps, and image size was 80*60 as shown in table I.

Tabel I: Parameters of Video dataset.

Video content	News, drama, sports etc.
Video length	Query video clips: 15s * 200 Video database sequence: 6 hours
Frame rate	14.98 fps
Frame number	Query video clips: 15s * 200 Video database sequence: 6 hours
Image format	PPM
Capture size	80*60

We captured 6 hours of video twice, one for video database sequence and the other for query video clips. Query video clips were generated by selecting video clips randomly for 200 times from the second video. Then we can perform search for 200 video clips from 6 hours of video. The threshold θ is 0.7 determined by preparing experiments according to FAR and FRR curve, which will be discussed later.

We utilized a 1-hour video sequence by selecting randomly from the second video to determine boundary threshold which were used to implement scalar VQ process (SVQ).

To suit the search task, quantization levels of HOG are set at 8 in θ -axis (totally 9) in the feature extraction stage. Thus, the number of histogram bins is total 512. Similarity calculation between the feature vector histograms will be quite faster compared with conventional algorithm which number of histogram bins is 4096.

4.1. Image Features of Conventional Algorithm

In conventional algorithm [1], they use small scaled images as video features. An image feature vector is defined as formula (6).

$$g(k) = (g_1(k), \cdots, g_j(k) \cdots, g_W(k)) \qquad (6)$$

where k is the frame number, j is the division number of the subimages, and W is the number of subimages. The $g_j(k)$ is the normalized intensity and is defined as formula (7), where $\overline{x}_j(k)$ is the average intensity in the j-th subsection.

$$g_j(k) = \frac{\overline{x}_j(k) - \min_i \overline{x}_i(k)}{\max_i \overline{x}_i(k) - \min_i \overline{x}_i(k)} \tag{7}$$
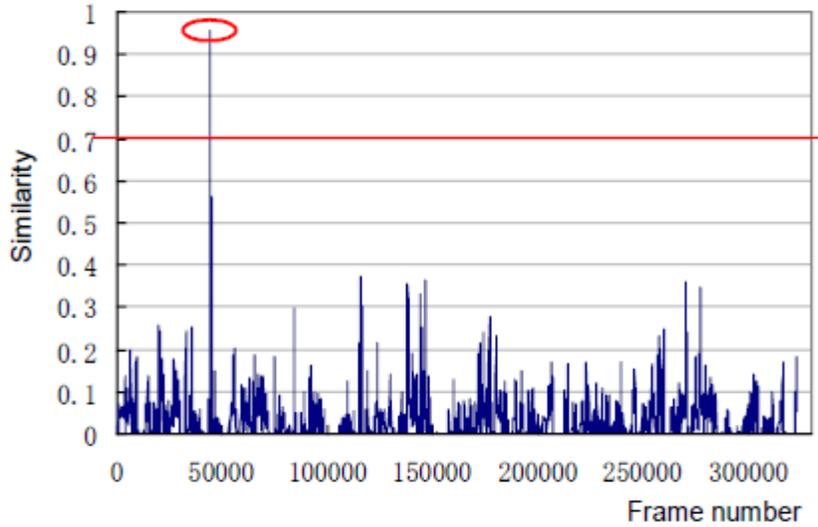
Figure 3. The example of search result.

4.2. Experimental Results

Figure 3 shows an example of result in 200 times search. The red ellipse marks the correct location of detected query video clip. Results of search accuracy are shown in figure 4. Search accuracy is shown as a function of window size. The perfect accuracy of 100% is obtained when window size is given 30sec. But even if window size decreases to 2sec, the accuracy still remains 98%.

We also compared our algorithm with the algorithm which does not utilize active search (full search), and conventional fast search algorithm described in section 4.1. Table II gives the approximate computational cost of the algorithms. As descried above, the number of histogram bins is total 512 in our proposed algorithm, 8 times smaller than that of conventional algorithm. From Table I, we can see the search time costs only 70ms, which is 271 times faster than full search, and also 6.7 times faster than the conventional fast search algorithm.

We also investigated the robustness of image features used in respective algorithms by adding Gaussian noise to the query video clips. Figure 5 shows how the search accuracy changes with the amount of noises. The curves with trigonal mark and foursquare mark stand for proposed algorithm and conventional algorithm, respectively. Our proposed algorithm achieves higher search accuracy than conventional algorithm. Although the search accuracy decreases with

increase the amount of Gaussian noises, it can be said that proposed algorithm is more robust for video search task than the conventional approach.

Table I: Approximate Computational Cost (CPU time).

Stage	Full search	Conventional	Proposed algorithm
Feature Extraction	540sec	540sec	550sec
VQ processing	50ms	50ms	35ms
Search	19sec	470ms	70ms
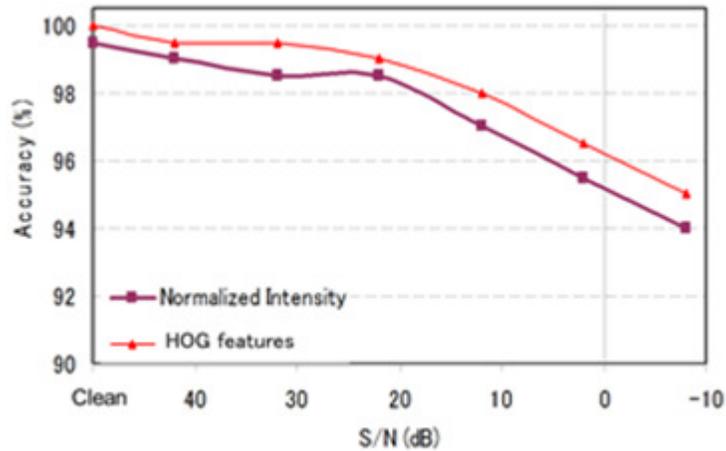
Figure 5. Gaussian noise vs search accuracy.
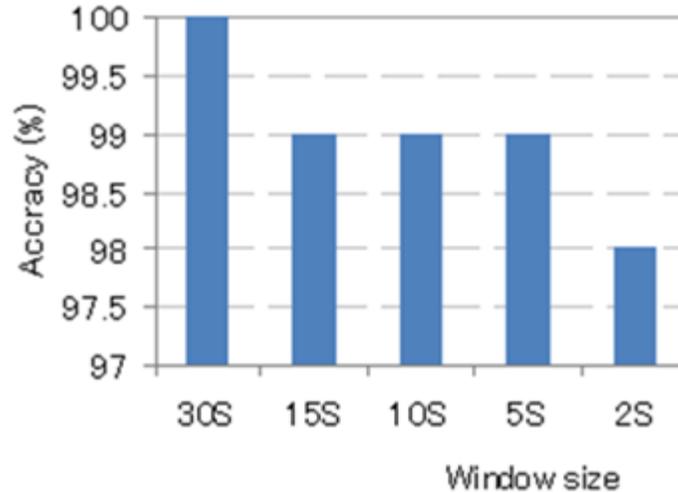
Figure 4. Window size vs. search

5. CONCLUSIONS

By using a new feature based on HOG based features, we present a fast and robust video search algorithm for video clips from large video database. The proposed search algorithm has been

evaluated by 6 hours of video to search for 200 video clips. Experimental results show that search time costs only 70ms, which is 271 times faster than full search, and also 6.7 times faster than the conventional fast search algorithm. Furthermore, proposed algorithm is more robust against Gaussian noise for video search task than the conventional approach.

REFERENCES

[1]   K., Kashino, T. Kurozumi, and H. Murase, "Quick AND/OR search for multimedia signals based on histogram features", IEICE Trans., J83-D-II(12), 2000, pp. 2735-2744.

[2]   S.S. Cheung and A. Zakhor, "Efficient video similarity measurement with video signature", IEEE Trans. on Circuits and System for Video Technology, 13(1), 2003, pp. 59-74.

[3]   A. Hampapur, K. Hyun, and R. Bolle, "Comparison of sequence matching techniques for video copy detection", SPIE. Storage and Retrieval for Media Databases 2002, 4676, San Jose, CA, USA, 2002, pp. 194-201.

[4]   V.V. Vinod, H. Murase, "Focused color intersection with efficient searching for object extraction", Pattern Recognition, 30(10), 1997, pp. 1787-1797.

[5]   N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection", In CVPR, pp. 886–893, 2005.

[6]   C. Hou, H. Ai, and S. Lao, "Multiview pedestrian detection based on vector boosting", In Asian Conference on Computer Vision, pp.18–22, 2007.

[7]   L. Agnihotre, N. Dimitrova, T. McGee, S. Jeannin, S. Schaffer, J. Nesvadba, "Evolvable visual commercial detector", In Proceedings of IEEE. International Conference on Computer Vision and Pattern Recognition, vol. 2, 2003, pp. 79-84.

[8]   R. Lienhart, C. Kuhmunch, W. Effelsberg, "On the detection and recognition of television commercials", In Proceedings of IEEE Conference on Multimedia Computing and Systems, 1997, pp. 509-516.

[9]   R. Mohan, "Video sequence matching", In Proceedings of the International Conference on Audio, Speech and Signal Processing, vol. 6, 1998, pp. 3679-3700.

[10] A. Araujo, M. Makar, V. Chandrasekhar, D. Chen, S. Tsai, H. Chen, R. Angst, and B. Girod, "Efficient video search using image queries", In Proceedings of International Conference on Image Processing, pp. 3082-3086, 2014.

[11] Q. Zhu, M. C. Yeh, K. T. Cheng, and S. Avidan, "Fast human detection using a cascade of histograms of oriented gradients" In CVPR, pp. 1491–1498, 2006.