FAST VIDEO RETRIEVAL VIA THE STATISTICS OF MOTION

*Jing-Fung Chen, Hong-Yuan Mark Liao and Chia-Wen Lin**

Institute of Information Science, Acadmeia Sinica, Taipei, Taiwan
Department of Computer Science and Information Engineering National, Chung Cheng University, Taiwan*

ABSTRACT

Due to the popularity of the Internet and the powerful computing capability of computers, efficient processing/retrieval of multimedia data has become an important issue. In this paper, we propose a fast video retrieval algorithm that bases its search core on the statistics of object motion. The algorithm starts with extracting object motions from a shot and then transform/quantize them into the form of probability distributions. By choosing the shot that has the largest entropy value among the constituent shots of an unknown query video clip, we execute the first stage video search. By comparing two shots with different lengths, their corresponding motion probability distributions are compared by a discrete Bhattacharyya distance which is designed to measure the similarity between any two distribution functions. In the second stage, we add an adjacent shot(either preceding or subsequent) to perform a finer comparison. Experimental results demonstrate that our fast video retrieval algorithm is powerful in terms of accuracy and efficiency.

1. INTRODUCTION

In the multimedia era, using an unknown video clip to retrieve a complete counterpart video in a video database will definitely be the future trend of our daily life. A piece of video clip may contain several consecutive shots and its first and last shots may be incomplete. In order for executing efficient video indexing/retrieval, many crucial technologies such as shot change detection [1, 2], shot representation [3, 4], key video frame/clip extraction [6], etc., have been developed in the past decade. In this paper, we shall use two gradual shot change detection algorithms [1, 2] developed by ourselves to detect the shot boundaries of an unknown video clip. Then, we compute the entropy of every constituent shot of the above mention video clip and pick the shot that has the largest entropy value as the first stage query. We extract object motions from this shot and then transform/quantize these local motions into the form of 2D probability distributions. The set of probability distributions extracted from the query shot is then compared with the probability distributions of the shots in the database. We choose the discrete Bhattacharyya distance that is designed specifically for comparing two distributions to execute the task. The above mentioned first-stage coarse search is able to significantly cut down the size of the search space. By choosing an adjacent shot of the first query shot, we concatenate the two shots to form the query of the second-stage coarse search. The motion feature used in the second-stage coarse search is the same as that used in the first stage. However, the order of the two consecutive shots is equivalent to the introduction of the "causality" effect. Experimental results demonstrate that our fast video retrieval algorithm is powerful in terms of accuracy and efficiency.

2. PREPROCESSING

Since a shot is the most primitive unit with semantic meaning that can be used for video retrieval, a powerful shot change detection algorithm is indispensable. In [1] and [2], we have proposed two powerful gradual shot change detection algorithms that can detect roughly about 80% of gradual shot transitions. In order to extract valid local motions from a shot, we used the above mentioned algorithms to process on several videos and obtained a large number of shots from these videos. For a complete shot, it is possible to cover several GOPs(group of pictures). Under the circumstances, the original design of an MPEG bitstream cannot be used directly for computing the local motion(object motion) between every consecutive anchor frame pair(anchor frame means I- and P- frames). For a P-frame in an MPEG bitstream, the forward motion vectors between itself and its reference frame can be directly extracted and used as a valid motion field. However it is not the case for an I-frame because an I-frame is actually intra-coded and containing no motion information inside it. For the above mentioned discontinuity problem that does exist between two consecutive GOPs, we propose to use the motion vectors of the B-frame which is located right after the last P-frame in each GOP to solve the problem. Suppose the forward motion vector and backward motion vector of a macroblock in the above mentioned B-frame are $\vec{F}$ and $\vec{B}$, respectively. Then $\vec{F}$ - $\vec{B}$ can be regarded as a pseudo motion vector between the last P-frame in the current GOP and the I-frame in the next GOP. The replacement can be represented graphically as showed

in Fig.1. Thus we can construct a motion vector field between any two anchor frames even the motion vector cannot be found in I-frames.

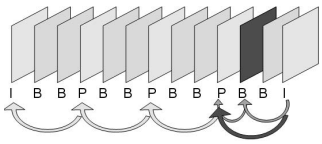

Fig. 1. Consecutive motion vector.

3. LOCAL MOTION EXTRACTION

In this study, the size of a video frame is 352×240, and therefore there are in total $22 \times 15 = 330$ macroblocks in a frame. However, we observed that the boundaries of a frame are relatively motionless and therefore should be discarded. In addition, the place in which the captions usually appear should also be discarded. Under these circumstances, we only use the central portion of a frame as valid macroblocks. The total number of macroblocks that are considered valid in our study is $20 \times 9 = 180$. Fig.2 is the illustration of where in a frame the valid macroblocks located.

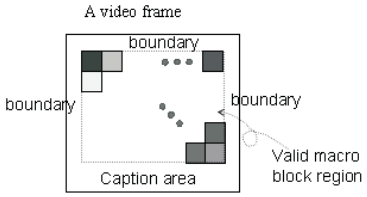

Fig. 2. The illustration of where in a frame the valid macroblocks located.

Now, we are ready to discuss how to calculate the statistics of motion from a valid macroblock sequence located in a shot. Using this statistics, we are able to conduct quantitative comparison between two distinct shots. The left hard side of Fig.3 illustrates a typical shot consisting of n anchor frames. From this shot, we can derive $n-1$ local motions between any two consecutive anchor frames. Since a local motion vector derived from two consecutive macroblocks in a shot sequence may be large in magnitude, we have to transform it into a smaller domain and, in the mean while, quantize it to facilitate the motion statistics calculation process. For a motion vector (x, y) located in the XY plane, we can transform it into the UV plane by the following equation:

$$\begin{aligned} u &= \lfloor x/I + 0.5 \rfloor \\ v &= \lfloor y/I + 0.5 \rfloor \end{aligned} \tag{1}$$

where I is an integer that can be used to control the degree of quantization(set I=5). The quantization procedure is able to group the motion vectors that are close to each other into one bin and makes the derived motion statistics more reasonable.

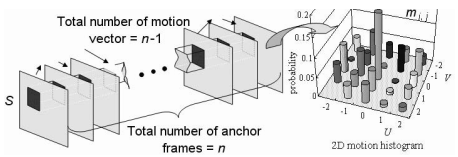

Fig. 3. An example showing how the motion vectors extracted from a macroblock sequence are projected onto the UV plane.

For calculating the statistics of motion from a valid macroblock sequence(as indicated in Fig.3), we have to do the following. First, let $m_{i,j}$ be the set of motion vectors of a valid macroblock sequence. The first macroblock of this macroblock sequence begins from the i-th row, j-th column of the valid macroblock region as indicated in Fig.2. The probability that the quantized(or transformed) motion vectors of this macroblock sequence falls into the bin (u, v) can be calculated as follows:

$$p(LM = (u,v) | LM \in m_{i,j}) = \frac{\#\{LM | LM = (u,v)\}}{L} \tag{2}$$

where LM represents a motion vector after transformation(or quantization) and L is the total number of quantized motion vectors in this valid macroblock sequence. $\#\{LM|LM = (u,v)\}$ means the number of quantized motion vectors that fall into the (u, v) bin. Fig.3 illustrates how to transform $n-1$ motion vectors into the normalized probability distribution map located in the UV plane. In the example, the range of U and V are both from -2 to +2. In addition to the probability distribution calculated above, we can also compute the entropy value of every valid macroblock sequence by the following equation:

$$\begin{aligned} H(S) = -\sum_{u,v} \{ & p(LM = (u,v) | LM \in m_{i,j}) \\ & \ln p(LM = (u,v) | LM \in m_{i,j}) \} \end{aligned} \tag{3}$$

The calculation of the entropy shown in Eq.(3) can be used to guide the selection among the constituent complete shots of an unknown query clip. Usually, a shot having the largest entropy value means the shot contains the most discriminating information that can be used to conduct an efficient search. In this work, we propose a two-stage process for efficient video retrieval. As described before, we shall adopt the complete shot that has the largest entropy value to execute the first stage search process. As to the second stage search process, we shall adopt a shot that is adjacent to the first chosen shot(either before or after) and make them work

together for more accurate search outcome. Fig.4 illustrates some possibilities about the selection of the second shot. The left hand side of Fig.4 shows a video clip(query) containing a number of shots. The first and the last shots, as indicated, are incomplete shots. Between the two incomplete shots, there are three complete shots. If the leftmost shot or the middle shot is chosen in the first stage search process due to its largest entropy(situation 1 or 2), then we use its subsequent shot as the chosen shot in the second stage search. On the other hand, if the rightmost shot(situation 3) is chosen for the first stage search, then its prior shot will be adopted to play the role in the second stage search process.

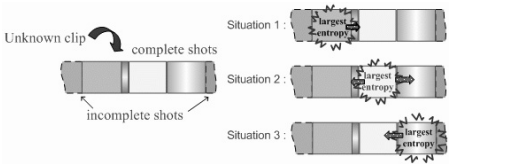

Fig. 4. The strategy of selecting the shot for the second stage search process.

For comparing two distinct shots, the comparison is of the form of comparing two probability distribution functions. Here, we shall use the so-called Bhattacharyya distance [5] to do the job. The Bhattacharyya distance is a well-known metric which is defined for measurement of the correlation between two arbitrary statistical distributions. For any two arbitrary distributions $p(x|\omega_1)$ and $p(x|\omega_2)$ of classes ω_1 and ω_2, respectively, the continuous form of the Bhattacharyya distance is defined as [5]:

$$D(\omega_1, \omega_2) = -\ln \int (p(\mathbf{x}|\omega_1)p(\mathbf{x}|\omega_2))^{1/2} d\mathbf{x} \qquad (4)$$

However, in the previous section we have described that the distribution of motion vectors in a shot is formulated as the discrete format. Therefore, we need a discrete Bhattacharyya distance to perform the shot comparison task. Let $m_{i,j}$ and $m'_{i,j}$ be the set of motion vectors extracted from the valid macroblock sequence at (i, j) location of two distinct shots(the two shots may have different lengths). Then, based on the definition made in Eq.(4), we can define a discrete Bhattacharyya distance for the above two distinct macroblock sequences as follows:

$$d(m_{i,j}, m'_{i,j}) = -\ln \sum_{u,v} \{p(LM=(u,v)|LM \in m_{i,j}) \; p(LM'=(u,v)|LM' \in m'_{i,j})\}^{1/2} \qquad (5)$$

For calculating the distance between two shots, we use the example shown in Fig.5 to explain how the Bhattacharyya distance works. Fig.5 illustrates two shots having different lengths. The number of anchor frames for the upper shot and the bottom shot are n and p, respectively($n \neq p$). Taking a valid macroblock sequence at the same position from the two different shots, we shall have $n-1$ and $p-1$ motion vectors, respectively, extracted from the upper shot and the bottom shot.

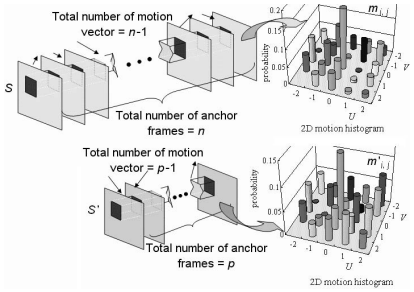

Fig. 5. An example showing how the motion vectors that belong to two macroblock sequences of two distinct shots are extracted and mapped onto the UV plane. The correlation between the two macroblock sequences is calculated by multiplying all bin-to-bin probabilities and then summing them up.

After transformation(quantization), we are able to quantize both motion vector sets onto the UV plane as indicated on the right hand side of Fig.5. The comparison using the Bhattacharyya distance shown in Eq.(5) can be carried out as follows. Eq.(5) is the Bhattacharyya distance designed to measure the similarity between two macroblock sequences located at the same position but from different shots. In order to calculate the overall Bhattacharyya distance between two arbitrary shots, one has to accumulate the measured distance from all macroblock sequence pairs located in the valid macroblock region as indicated in Fig.2. The equation for calculating the overall similarity, $D(S, S')$ is as follows:

$$D(S, S') = \frac{\sum_{i,j} d(m_{i,j}, m'_{i,j})}{N} \qquad (6)$$

where S and S' represent two distinct shots, and N represents the total number of valid macroblock sequences existing in a shot.

4. EXPERIMENTAL RESULTS

In order to show the effectiveness of the proposed method, we have tested our algorithm against a 1682-shot video database. First, we used the gradual shot change detection algorithms proposed in [1, 2] to extract 1682 complete shots from six digital videos. For each constituent shot in the database, we used the same method as described in this paper to calculate its statistics of motion(off-line). The length of the six digital videos were 55 minutes(503 shots, documentary, video#1), 52 minutes(405 shots, documentary, video#2), 29

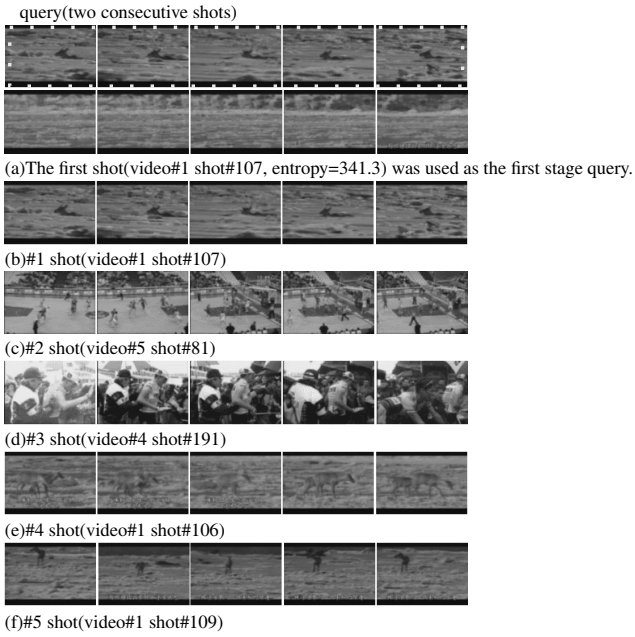

Fig. 6. Retrieved results of the first-stage coarse search. (a) the query shot was video#1 shot#107; (b)-(f) the top five retrieved shots out of the 1682-shot database.

minutes(241 shots, commercial, video#3), 38 minutes(193 shots, news, video#4), 38 minutes(283 shots, sports news, video#5), 17 minutes(57 shots, home video, video#6), respectively. The reason that we chose these video was due to their variety. In order to test the accuracy of the method, we used a video clip that was port of the test videos as the query. Based on the entropy value, we chose shot#107 of video#1(Fig.6(a)) as the query shot of the first stage coarse search. The top 5 retrieved results were shown in Fig.6(b)-(f). It is obvious that the retrieved results based on the statistics of motion was very efficient, but not necessary very accurate. In the second stage coarse search, we added shot#108 of video#1(Fig.7(a)) into the comparison process. Since the causality factor as well as the motion statistics of one more shot were added to enhance the power of the feature set, the top 4 retrieved results(Fig.7(b)-(e)) were all very close to the query shot pair. By using only two coarse search stages, we were able to retrieve a very close data sets. In the future, we shall put our emphasis on introducing a new comparison strategy that bases its search on finer structural features.

5. CONCLUSIONS

We have proposed a fast video retrieval algorithm which could retrieve video simple clip efficiently and accurately. The proposed approach used the statistics of motion extracted from a shot as the search features, and we used entropy to decide the order of query shots. The proposed algorithm could compare two arbitrary simple clips with different lengths by a so-called Bhattacharyya distance and returned the accurate result in mini-seconds. Experimental results have demonstrated that the proposed fast video retrieval algorithm is indeed powerful.

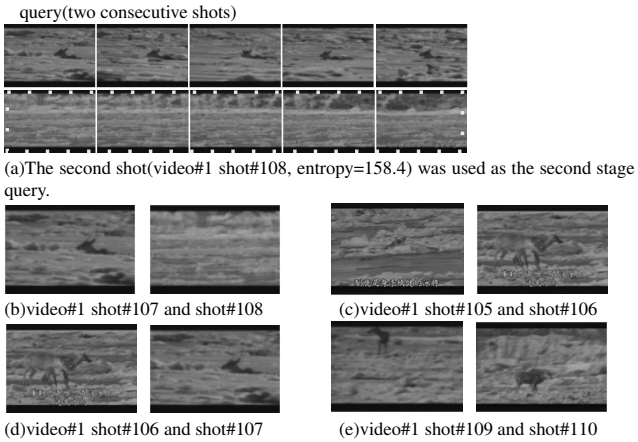

Fig. 7. Retrieved results of the second-stage coarse search. (a) the second query shot; (b)-(e) the top four retrieved shot pairs.

6. REFERENCES

[1] C. W. Su, H. R. Tyan, H. Y. Mark Liao and, L. H. Chen. A Motion-tolerant Dissolve Detection Algorithm. *Proc. IEEE Int. Conf. on Multimedia and Expo, Lausanne, Switzerland*, 26–29, August 2002.

[2] C. C. Shih, H. R. Tyan and, H. Y. Mark Liao. Shot Change Detection based on the Reynolds Transport Theorem. *Proc. IEEE 2nd Pacific-Rim Conference on Multimedia, Beijing, China, Lecture Notes in Computer Science*, 2195:819–824, October 2001.

[3] R. Fablet and, P. Bouthemy. Statistical motion based object indexing using optic flow field. *15th IAPR, ICPR*, 4:3–7, September 2000.

[4] Y. F. Ma and, H. J. Zhang. A New Perceived Motion Based Shot Content Represention. *Int. Conf. on Image Processing Proceedings*, 3:7–10, October 2001.

[5] L. F. Chen, H. Y Mark Liao, J. C. Lin and, C. C. Han. Why Recognition in a Statistics-Based Face Recognition System should be Based on the Pure Face Portion: a Probabilistic Decision-Based Proof. *Pattern Recognition*, 34(5):1393–1403, 2001.

[6] A. K. Jain, A. Vailaya and, W. Xiong. Query by Video Clip. *Multimedia System*, 7:369–384, 1999.