

GROUP-OF-FRAME/PICTURE COLOR HISTOGRAM DESCRIPTORS FOR MULTIMEDIA APPLICATIONS*

A. Mufit Ferman[§], S. Krishnamachari[†], A. Murat Tekalp[§], M. Abdel-Mottaleb[†], and R. Mehrotra^{*}

[§]Dept. of Electrical Engineering and Center for Electronic Imaging Systems
University of Rochester, Rochester, NY 14627

[†]Philips Research USA
345 Scarborough Road, Briarcliff Manor, NY 10510

^{*}Imaging Science Technology Lab
Eastman Kodak Company, Rochester, NY 14650

ABSTRACT

Joint representation of color-based features for multiple images or a video segment is an important task for visual information management systems. Generally, a key-frame or key-image is selected from such a group, and the color-related features of the entire collection are represented with those of the chosen sample. Such methods are highly dependent on the quality of the representative sample, and may lead to unreliable results. We present a set of histogram-based descriptors that reliably capture the color content of multiple images or video frames. These descriptors are defined for a group-of-frames (GoF) or a group-of-pictures (GoP). A single representation for the entire collection is obtained by combining individual frame or image histograms in various ways. We demonstrate the efficacy of GoF-histograms for video segment retrieval and the GoP-histograms for fast image search. This descriptor has been accepted to the Working Draft of MPEG-7, the evolving ISO standard for multimedia content description.

1. INTRODUCTION

As the size of multimedia databases increases, it becomes necessary to represent multiple frames or images with efficient descriptors. A common method for joint representation of color-based features is to select one or more representative images or frames from a group, and to represent the color properties of the entire collection with those of the chosen sample(s). Such methods are highly dependent on the quality of the representative images, and may lead to unreliable results when the selections are made inconsistently.

In this paper we present a set of histogram-based descriptors that reliably capture and represent the color properties of multiple images or video frames. These descriptors are defined for a *group-of-frames* (GoF) or a *group-of-pictures* (GoP), and combine individual frame/image histograms in various ways to generate a single representative histogram for the entire collection. We outline some application

areas of the GoF/GoP color histogram descriptors for visual content management in image and video databases. In particular, we demonstrate the efficacy of the GoF histograms for video segment retrieval, and illustrate the usefulness of the GoP histogram to characterize a collection of similar images for applications such as fast query-by-example (QBE)-based image search.

The GoF/GoP histogram descriptors have recently been proposed to the MPEG-7 standard as a color descriptor and have been accepted into the working draft after a series of vigorous evaluation experiments. MPEG-7 (formally "Multimedia Content Description Interface") is an evolving ISO standard that aims to specify a standard set of *descriptors* to describe various types of multimedia information, *description schemes* to define the structure and semantics of these descriptors, and a *description definition language* to express the multimedia content description [1]. Feature extraction and search engines form the non-normative part of the standard. This approach will enable development of applications that will utilize the MPEG-7 descriptions without specific ties to a single content provider.

2. GOF/GOP HISTOGRAM DESCRIPTORS

In this section we outline the definitions of the various GoF/GoP histogram descriptors; namely, average, median and intersection histograms. One or more of these descriptors can be used to represent the color content of a single GoF/GoP.

2.1. Average Histogram

The average histogram is computed by accumulating the individual frame histograms, and then normalizing each bin by the number of frames in the GoF. A potential problem with using sample averages to compute the GoF histogram is the sensitivity of the mean operator to outliers. Given M samples, any data point has a weight of M^{-1} in the computation of the

* To appear in the Proceedings of International Conference on Image Processing - 2000, Vancouver, Sep 2000.

mean [2]; hence a deviant data value can lead to a biased sample average that is not representative of the entire set.

2.2. Median Histogram

An alternative estimator to the sample mean is the median, which can efficiently eliminate the outliers present in the data. To compute the value in every bin in the median histogram, the ascending list of frame histogram values is constructed for the duration of the GoF, and the median of this list is determined. One concern regarding the use of the median histogram is the increased computational complexity, since sorting needs to be performed for each histogram bin. Using a common algorithm like quicksort, the total number of comparisons required to sort the histograms of a GoF with M frames is $B \times O(M \log M)$, where B is the number of bins.

Average and median histograms can be viewed as members of a family of representative histograms we call *alpha-trimmed average histograms*. This set of histograms is generated using the *trimmed mean* operator [2], which works by sorting the sample points (in this case, the array of frame histogram values for each bin) in ascending order, and averaging only the central members of the ordered array.

2.3. Intersection Histogram

Histogram intersection is a popular scalar-valued similarity measure for color-based indexing and object recognition [3]. It yields the number of pixels that have the same color in two images. In contrast, the intersection histogram [4] that we propose is itself a histogram, computed over the range of frames in a GoF. The value of the j th bin in the intersection histogram $IntHist_k$ of the k th GoF is determined by

$$IntHist_k(j) = \min\{H_i(j)\}, i = 1, \dots, \text{no. of frames} \quad (1)$$

Each bin value in the intersection histogram thus represents the number of pixels of a particular color that appear in all of the GoF frames.

The intersection histogram is characteristically different from the average and median histograms, in that it provides the “least common” color traits of the given GoF, rather than an estimate of the color distribution. This property of the intersection histogram makes it appropriate for fast identification of the GoF in which a query image occurs as shown in Section 3.2

3. GOF HISTOGRAM APPLICATIONS

The GoF color histograms present a versatile tool for numerous applications in video content management, including QBE-based retrieval applications, automatic key frame selection, and shot grouping/semantic scene reconstruction. In the rest of this section, we present a detailed description of experimental results obtained for video segment retrieval and GoF identification with intersection histograms.

The experiments were conducted on 1544 GoFs extracted from eight video (~180 minutes of video) picked out of the MPEG-7 content set. The selected video data represents a heterogeneous set drawn from a wide range of content domains.

Each identified GoF is a shot, a dissolve, a fade, or a wipe; blank frames and segments where the actual segment boundaries were hard to determine were excluded from the test set. The three types of descriptor histograms were then computed for every GoF. Additionally, the histogram of a key frame was identified for each GoF in order to evaluate the performance of key frame-based video segment retrieval methods. The key frame selection method involves an exhaustive search over all GoF frames to determine the *optimal* frame that minimizes an error criterion. All histogram computations were carried out in the YCbCr color space, with the luminance component uniformly quantized to 8 bins, and the chrominance channels to 12 non-uniformly spaced bins each, yielding a total of 1152 bins for every histogram. A total of 31 query GoFs and the corresponding ground truth data were then picked from the GoF database to be used in the experiments. The queries range from almost identical static GoFs to dynamic scenes and edit effects/transitions. Histogram similarities were computed using the L_1 , L_2 , χ^2 and *histogram intersection* comparison measures.

3.1. Video Segment-to-Segment Matching

The performance measure used to evaluate the retrieval results is the average normalized modified retrieval rank (*ANMRR*), which has been employed in MPEG-7 standardization activities as the primary evaluation criteria [1]. Given a query set and the corresponding ground truth data, *ANMRR* is designed to determine how many of the correct GoFs are retrieved, how highly they are ranked among the retrieved items, and how many relevant GoFs are missed. *ANMRR* values range between [0,1]; a low value of *ANMRR* denotes a high retrieval rate (small number of misses) with relevant items ranked at the top. On the other hand, *ANMRR* = 1 represents the worst case, with none of the relevant items in the database present in the top retrievals. Details on the *ANMRR* computations are presented in the Appendix.

The GoF histogram-based video segment retrieval results are shown in Table 1. Retrieval experiments were carried out using average, median and the selected key frame histograms. It is observed that *ANMRR* values are the smallest for *AvgHist* and *MedHist*; both GoF histograms consistently outperform the optimal key frame histogram. When individual queries are considered, it is observed that certain GoF histogram types are more appropriate for retrieval of specific types of GoFs. Generally, it is advantageous to use the median histogram in cases of occlusion by a large object, or when gradual variations are observed in the background and/or luminance characteristics. Overall, the performances of the average, median, and key frame histograms are on par, especially in cases where (a) camera remains stationary and object movement is not too significant; and (b) there is object/camera movement, but the scene is dominated by the background object.

3.2. Fast GoF Search with Intersection Histograms

A significant application area for intersection histograms is fast identification of the particular GoF that a query image belongs to. The method is especially useful for large databases, since

unlike regular similarity-based search methods it retrieves only a small number of candidate items from the database, and the correct GoF is guaranteed to be among the retrievals (if the query frame does in fact appear in the database). The search method can be very useful in large stock footage, sports, and news archives, as well as home video collections, where an individual frame is frequently captured from a video stream and utilized as a still frame in printing, publishing, etc.

Given a query frame f , the bin-wise differences between the frame histogram H_f and the intersection histogram of the k th GoF must, by definition, be non-negative for all bins if $f \in k$. The fast search method involves computation of bin-wise differences between H_f and the intersection histogram $IntHist_k$ of each GoF k in the database:

$$D_j = H_f(j) - IntHist_k(j), \quad j = 1, \dots, B, \quad (2)$$

with B the number of bins in each histogram. If $D_j < 0$ for any bin, the GoF is immediately rejected as a candidate. Otherwise, a *match coefficient* $C_{f,k}$ is computed as

$$C_{f,k} = \frac{1}{N_f} \sum_j D_j, \quad (3)$$

where N_f is the total number of pixels in frame f . The $C_{f,k}$ values are then rank-ordered to yield a list of GoFs to which the query frame may belong. We express the performance of the proposed search method quantitatively by the average number of retrieved items per query, and the ranking of the correct GoF among the retrievals.

The proposed method was tested using the key frames picked from the 31 query GoFs and the search was performed over the 1154 GoFs. Average number of candidate GoFs returned per query was 16; in 28 of the 31 queries, the correct GoF was retrieved as the top-ranked candidate. For the other three cases, the correct GoF was ranked 2nd (in 5 retrieved items), 3rd (out of 12 retrieved items), and, in the worst case, 6th (out of 20 retrieved items).

4. GOP HISTOGRAM APPLICATIONS

In a typical image collection, many images tend to be similar because they are captured with the same background or at the same scene. Furthermore, users are often more interested in finding a collection of similar images than a single image. Hence, grouping similar images and characterizing them by a single color descriptor can be useful in (a) expediting user access in QBE applications, and (b) browsing non-linearly through a large database to identify images of interest. In the following section we present one of the potential applications, namely, fast search over an image database. We demonstrate that by using GoP-histograms, it is possible to perform a fast search over a large database without significant degradation in retrieval accuracy.

4.1 Fast Search using GoP-Histogram Descriptor

The first step in using GoP-histograms is to cluster the given collection of images into different groups based on their

similarity. This requires a similarity measure with which to compare images, and a clustering technique for grouping the images using the specified measure. We have chosen the *histogram intersection* as the similarity measure, and the *hierarchical clustering* algorithm [5] as the clustering method. For a database of n images, first the similarity between every image pair is computed. Each image is then assigned to a single cluster, yielding n initial clusters. These clusters are iteratively merged based on their similarity. The algorithm terminates when the number of clusters is reduced to a pre-defined number c , and yields a tree structure for each cluster [6].

Once the set of clusters for the image database is obtained, it is necessary to obtain a representation for each cluster. This representation can be any of the group histograms presented in Section 2, *i.e.*, the average, median, or intersection histograms of the images in the cluster.

To demonstrate the usefulness of the GoP-histogram descriptor, we performed QBE search using the GoP-histograms of the clustered image groups. The search strategy is as follows: the query image is initially compared with the GoP-histograms of all the clusters. This yields the subset of clusters with the largest similarity to the query image. Subsequently, all the images (histograms) in these clusters are compared with the query image. The results of retrieval and the computational requirement using the above method are compared against an exhaustive search, where the query image is individually compared against all the images in the database without any clustering.

We define the following measure to quantitatively compare the retrieval results obtained using the proposed approach to those for exhaustive search: Let us assume that the user is interested in only the top K best matches. Let M is the number of images present in the top K results returned by both the clustering based and the exhaustive search methods. The retrieval accuracy of the fast search method is defined as $100 * M / K$. This ratio is averaged over multiple queries to obtain the average retrieval accuracy.

A database of 3856 images, selected from two collections of Corel Professional Photo CD-ROMs and the California Department of Water Resources image database, were used in the experiments. The images were clustered into 175 groups using the algorithm above, and the three different types GoP-histograms were computed for each group. Cluster sizes ranged from 5 images for the smallest group to 80 for the largest one. A set of 300 images that are not a part of the database were used as query images. The query images were initially compared with the GoP-histograms of all groups, and the L top-ranked groups were identified. All images in these selected groups were then compared exhaustively with the query image to determine the most similar images. The average retrieval accuracy and the average number of required similarity comparisons were computed by averaging over all 300 queries. The experiment was repeated for $L = \{2, 3, 6, 9, 12, 18, 24, 30, 40, 50\}$. Figure 2 shows the plot of the average retrieval accuracy against the average number of similarity comparison required for the three type of GoP-histograms. Each plot contains 8 points, corresponding to each distinct value of L . From the figure it can be seen for $K=20$, a 90% retrieval accuracy can be obtained with about 400 similarity comparisons

for the average GoF-histogram case, as opposed to 3856 comparisons that would be required with exhaustive search. This is equivalent to a speed-up by a factor of about 10. With the median histogram 90% accuracy can be obtained with about 1200 similarity comparisons, a speed-up factor of a little over 3. The results show that the intersection histogram is not suitable for this specific application.

5. CONCLUSIONS

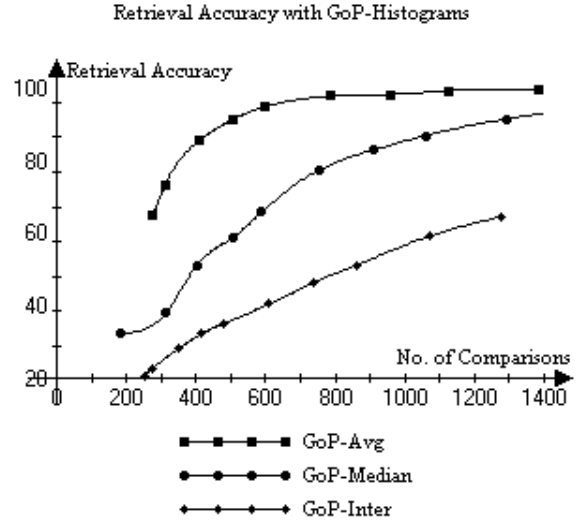
In this paper we have presented various types of histogram-based descriptors for joint color content representation of multiple images and video frames. The family of alpha-trimmed average histograms, which include the average and median histograms as special cases, provides a robust set of color descriptors that can eliminate the effects of aberrant frames within a GoF. This set of descriptors consistently outperforms key frame-based representations for video segment retrieval applications. The intersection histogram, on the other hand, reflects the number of pixels of a given color that is common to all the frames in the GoF, and can be employed in specific applications to yield efficient and reliable results. The fast search application for image databases demonstrates that considerable computational gain (speed-up factor of about 10) can be obtained using appropriate clustering and the GoP-histograms. The proposed descriptors are also appropriate for other tasks within a multimedia database management system, such as aggregation of GoFs for semantic scene reconstruction, and representative frame selection for visual summarization. These applications are currently under investigation.

6. REFERENCES

- [1] "MPEG-7 Visual part of eXperimentation Model Version 2.0," MPEG-7 Output Document ISO/MPEG, Dec 1999.
- [2] J. B. Bednar and T. L. Watt, "Alpha-trimmed means and their relationship to median filters," *IEEE Trans. Acoustics, Speech, and Signal Proc.*, vol. ASSP-32, no. 1:145--153, 1984.
- [3] M. J. Swain and D. H. Ballard, "Color indexing", *Intl. J. of Comp. Vision*, 7(11):11--32, 1991.
- [4] A. M. Ferman and A. M. Tekalp, "Multiscale content extraction and representation for video indexing," *Proc. Multimedia Storage and Archiving Systems II*, vol. SPIE 3229, Dallas, Nov. 1997.
- [5] A. K. Jain, *Fundamentals of Digital Image Processing*, Prentice Hall, 1986.
- [6] S. Krishnamachari and M. Abdel-Mottaleb, "Hierarchical clustering for fast image retrieval", *Proc. Storage and Retrieval for Image and Video Databases VIII*, pp 427-435, Jan. 1999.

	L_1	L_2	χ^2	HI
<i>AvgHist</i>	0.041367	0.089982	0.038833	0.051791
<i>MedHist</i>	0.042614	0.090640	0.044467	0.065469
<i>Opt. Kfr</i>	0.053500	0.101852	0.045312	0.062298

Table 1. ANMRR values for various GoF histograms and distance measures. The optimal key frame histograms have



been selected using the L_1 distance norm as the error criterion.

Figure 1. Plot of percentage retrieval accuracy against the number of required similarity comparisons for different types of GoP-histograms

APPENDIX

The details of the ANMRR computations are presented here. First a set of query images and the corresponding ground truth images that are considered similar to the query are selected manually. Let the number of ground truth images for query q be $NG(q)$. Let $K = \min(4 \times NG(q), 2 \times GTM)$ where $GTM = \max\{NG(q)\}$ for all queries. For each ground truth image k retrieved in the top K retrievals, compute the rank, $Rank(k)$, of the image, counting the rank of the first retrieved item as one. A rank of $(K+1)$ is assigned to each ground truth image that is not in the first K retrievals. Compute the *modified retrieval rank* $MRR(q)$ for query q as follows:

$$MRR(q) = \sum_{k=1}^{NG(q)} \frac{Rank(k)}{NG(q)} - 0.5 - \frac{NG(q)}{2}$$

The *normalized modified retrieval rank* ($NMRR$) is then computed as

$$NMRR(q) = \frac{MRR(q)}{K + 0.5 - 0.5 * NG(q)}$$

Note that $NMRR(q)$ will always be in the range of $[0.0, 1.0]$. Finally, the average of all $NMRR$ values is computed over all queries to yield the ANMRR.