Example-Based Color Vehicle Retrieval for Surveillance

Lisa M. Brown

IBM T.J. Watson Research Center
Hawthorne, NY 10532
lisabr@us.ibm.com

Abstract

In this paper, we evaluate several low dimensional color features for object retrieval in surveillance video. Previous work in object retrieval in surveillance has been hampered by issues in low resolution, poor segmentation, pose and lighting variations and the cost of retrieval. To overcome these difficulties, we restrict our analysis to alarm-based vehicle detection and as a consequence, we restrict both pose and lighting variations. In addition, we study the utility of example-based retrieval to avoid the limitations of strict color classification. Finally, since we perform our evaluation at run-time for alarm-based detection, we do not need to index into a large database. We evaluate the efficiency and effectiveness of several color features including standard color histograms, weighted color histograms, variable bin size color histograms and color correlograms. Results show color correlogram to have the best performance for our datasets.

1. Introduction

The problem we address is vehicle retrieval based on color in video surveillance. As part of a larger effort to provide cutting edge video surveillance in urban environments, we are interested in providing automatic event detection and extensive search capabilities [1]. In our prior work [2], we described a method to perform color classification of people and vehicles for video surveillance. Objects are first segmented based on background subtraction and tracking. Objects are then classified as one of six colors based on their histogram in quantized HSL space. The primary color of the object is determined using a rule-based approach.

Three problems arise from this type of color classification. The first is the issue of color constancy. Although people perceive an object to be the same color across a wide range of illumination conditions, the actual pixels of an object may have values which range across the color spectrum depending on the lighting conditions and relative pose. Secondly, it is difficult to accurately segment moving objects from the background. Shadows are often part of the object and errors exist in the segmentation due to the similarity of the object of interest, other objects in the scene and the background model. Lastly, complex objects are not predominately one color. Certain aspects of objects are of interest to the human and these depend on the type of object and application.

To overcome these shortcomings and to study the underlying issues, we limit the scope of the problem to objects which are vehicles. We also modify the problem in two key ways and at the same time create a new functionality in our surveillance system. First of all, rather than performing color classificaiton, we perform example-based color search. This eliminates the problem of defining an object into a single color class. Also, to the extent that the query and retrieved events occur under similar conditions, the example-based approach has the potential to reduce the issues due to lighting variations.

The second modifcation we introduce is to retrieve objects in real-time from a specific alarm-based set of events rather than from the set of all tracked objects stored in database. This greatly decreases the errors based on segmentation and tracking (since not all objects need to tracked) while minimizing the variations due to pose. More importantly, this eliminates the need to index into a large database. Instead, we test the query example against each alarm at run-time to determine the matching score. Events can then be retrieved in rank order.

Finally, this framework can be used to determine the feature space and similarity metric most effective for more standard color retrieval in surveillance applications. In this way, the necessary storage and indexing issues can be addressed independently from the issues of color discernment.

2. Related Work

Prior work in color-based retrieval is primarily from the field of content-based image retrieval (CBIR) and aims to search image databases for specific images that are similar to a given query image based on matching features

978-0-7695-4264-5/10 $26.00 © 2010 IEEE
DOI 10.1109/AVSS.2010.59

derived from the image content [3]. However, in a typical surveillance application, the user is not interested in a set of images but rather a set of objects of interest. In our case, this set is specifically the set of vehicles traversing the scene.

Additionally, in typical multimedia applications, such as [4,5] the subject matter of the images/videos is very broad. Unlike these systems which deal with a diverse and semantically varied set of images, we are focused on distinguishing very similar image objects: vehicles as seen by the same camera at the same location under poor lighting conditions and at relatively low resolution.

Since color is in many ways the most significant and distinguishing visual feature for retrieval, there has been a great deal of research devoted to developing color features for search. Color histograms are widely used because they are robust to occlusions and lighting and view changes. The pioneering work of Swain [6] showed the effectiveness of color histograms to distinguish a large number of objects and introduced a fast distance metric (incremental histogram intersection) to enable real-time indexing into a large database.

Investigators have since explored a wide range of color spaces beyond RGB for color representation, including normalized or transformed RGB [7], perceptually uniform spaces such as HSV [8], and L*u*v [3], and opponent color spaces such as Lab and YcbCr [9,10]. In order to avoid instability in HSV along the gray axis, a weighting system was developed by [11], and a nonuniform binning system by [12]. Several investigators have concluded that the HSV based features provide the best performance [8,13].

An important limitation of color histograms as a feature is their independence to spatial relationships. On the one hand, as a global measure, histograms are advantageous since they are invariant to many spatial transformations. On the other hand, they cannot distinguish between many images which are fundamentally different. To address this problem, researchers have developed many spatial color features including: compact color moments [12], partitioned images such as region-based color and binary sets [3,14], histogram refinement (local spatial properties in the histogram called color coherence vectors[15]) and color correlograms[16]. The most successful of these approaches has been the color correlograms. [17].

3. Feature Analysis

We examined several alternative color features for our query-based vehicle detection. The first feature is the basic RGB histogram with (NxNxN = N^3) bins. The RGB histogram can be made more invariant to lighting changes by normalizing the pixel value distributions.

$$\begin{pmatrix} R' \\ G' \\ B' \end{pmatrix} = \begin{pmatrix} \dfrac{R - \mu_R}{\sigma_R} \\ \dfrac{G - \mu_G}{\sigma_G} \\ \dfrac{B - \mu_B}{\sigma_B} \end{pmatrix}$$

Alternatively, we use the HSV color space. This space has several advantages. Since it is perceptually uniform, the distance between colors are more meaningful. Furthermore, it is straightforward to map the samples to human perceived “culture” colors. However, as mentioned in the related work section, the hue becomes unstable around the gray axis. There are two approaches to make this more robust. The first is to weigh each sample inversly proportionally to saturation so that as a sample become progressively more “gray” the color information is reduced [11].

The second approach is to create non-uniform bin sizes so that pixels which cannot be distinguished are not arbitrarily put in different bins [12]. Note that intensity information has a different value when indexing since it is not invariant to lighting conditions. However, many of the vehicle images we are testing do not have any significant color information.

A third approach, which we call “color split” is based on splitting the HSV space into chromatic and achromatic information. The approach has two stages. In the first stage, images are classified as either containing significant color informaion or not. In the second stage, gray histograms are computed for achromatic images and hue histograms are computed for chromatic images. This approach is based on segmenting HSV space along an intensity/saturation curve. This is explained in more detail in [2].

Figure 1 shows the results of splitting the chromatic vs achromatic along different curves. The top figure (left) shows the original image. The middle figure shows the pixels with hue (shown in white) for one curve. The right figure shows the pixels with hue along a curve closer to the central gray axis. In the right figure, it can be seen that pixels which are part of the road are classified as chromatic (perhaps because of some color in the lighting or reflections.) The color of these pixels are not meaningful and should not be included in the histogram. The two histogram below correspond to the hue pixels in the middle and right images respectively. The true color peak (in this case, red due to the color of the car) is difficult to ascertain from the second histogram because of the noise due to the pixels near the gray axis.

The last approach used is based on the color correlograms [16]. Color correlogram can be used with different color spaces and different number of distances between pixels. Altogether, we have six methods: RGB, transformed RGB, weighted HSV, non-uniformed bin HSV, color split, and color correlogram.

The chart below summarizes the methods and parameter settings used for the evaluation.

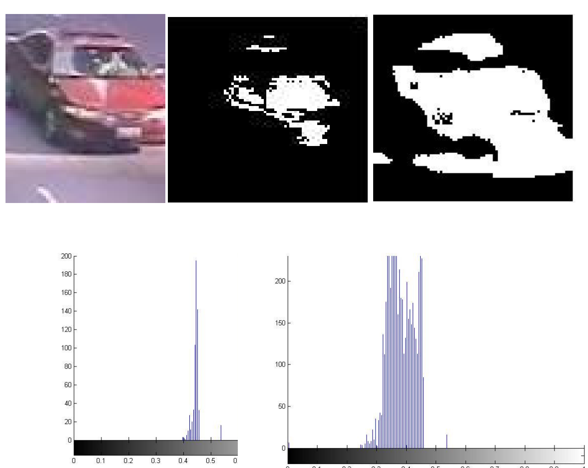
RGB (8x8x8) = 512 bins
Transformed RGB (8x8x8) = 512 bins, normalization for each image
Weighted HSV Histogram (Van der Weijer)
- (8x8x8) = 512 bins
- Weights based on inverse saturation

Weighted Variable Bin Size (Lei)
- 36 bin HSV
- 8 bins gray (low sat/int regions I & II)
- 4x7 bins color (hi sat region III)
- Gray: low weight, med sat/int: med weight, hi sat: hi weight, weights = (.5, 1.0, 2.0)

Color Split – Segment Color and Achromatic,
- Based on threshold (20%, tuned)
- Test against same class (color/achromatic)
- 15 bins (7 color, 8 gray)

Color Correlogram (Huang)
- (1) 15 bins HSV Quantized (7 color, 8 gray)
- (2) 36 bin Weighted Variable-Bin Size
- (3) 512 bin Transformed RGB
- distance 0,1,2 & 3 pixels apart

Figure 1. Left: Original image. Middle: Chromatic Segmentation with conservative saturation/intensity curve. Right: Chromatic segmentation with low saturation/intensity curve. Below corresponding hue histogram for middle and right segmentation.



3.1. Distance Metrics

To compute the ranking for each image in the database against the query image, a distance metric is needed to compute the similarity between examples. Metrics are typically L_1 or L_2. L_1 metrics include absolute difference or histogram intersection (made faster by incremental method of Swain [6]). Here, we use histogram intersection. Other metrics include the average color distance and a quadratic measures [18]. Sebe [8] introduces a new metric based on a maximum likelihood assuming we have classified training data. In practise, L_1 performs better than L_2 because it is more robust to outliers. It is further improved by weighting the absolute difference by the sum of the size of the histogram elements so that the metric is a relative difference measure. For color correlogram, we used absolute difference. The normalization was used in all cases.

4. Experimental results

Our objective is to determine the feature type which is best able to retrieve vehicles of similar coloring based on the query image. However, in order to have an initial set of ground truth data for evaluation, we first coarsely classified vehicles in five primary color categories: dark (dark silver & black), light (light silver & white), red, yellow and blue. We then tested each of the 6 color feature types for two data sets for their ability to rank vehicles of the same color class as high as possible.

4.1. Data Collection, Ground Truth & Metrics

Data was collected using several 4 hour sequences from urban traffic scenes. Only clean views of unobstructed vehicles were captured based on the detection of moving objects in a region of interest with an acceptable size, aspect ratio and direction of motion. Each captured image contains an associated image mask indicating the pixels which are in the foreground.

Figure 2. Examples of vehicle images from Dataset II. Dark, light, blue, red, and yellow images are shown in each of the five columns. Examples of the size variation can also been seen.

Two datasets were used for evaluation. The first set (Dataset I) contains 64 images of vehicles in varying lighting conditions in 5 solid colors: white(21), black,(13) blue(6), red(10), & yellow(14). The second set (Dataset II) contains 104 images with 3 blue, 30 dark, 62 light, 5 red, and 4 yellow. This set was obtained using all results which satisfied the collection criteria from a 1 hour sequence. The distribution of colors is therefore fairly representative of the typical distribution. The minimum resolution image was (41x93), the maximum (75x181) and the average (54.3x123.9).

Two performance measures were computed:

Within 100% percent of results in retrieval set that are same color where retrieval set is size of perfect results

Within 120% percent of results in retrieval set that are same color where retrieval set is 20% larger than perfect results

4.2. Results

Results on each dataset are shown in Tables 1 & 2.

Table 1: *Comparison of different methods on Dataset I*

Method	*Within 100%*	*Within 120%*
Weighted	71%	76%
RGB	78%	82%
Color Split	80%	82%
Transformed RGB	83%	85%
Weighted Variable Bin Size	82%	86%
Transf. RGB Color Correlogram	86%	90%
Variable Bin Size Color Correlogram	88%	92%
15- bin HSV Color Correlogram	89%	100%

The best results on Dataset I were obtained with the color correlogram. We decided to test the color correlogram with various color spaces to see if we could get further improvements. The best color correlogram result was obtained with the 15 bin HSV quantization (8 gray bins, 7 color bins,) where color/gray split was performed using the conservative saturation/intensity curve.

The best results on Dataset II were also obtained with the color correlogram. However, in this set, the transformed RGB color space seemed slightly better. We believe this is because in this set the majority of the test cases are gray (black or white) vehicles and therefore the normalization peformed by the transformed RGB quantization was more important. Notice that this was the more difficult dataset.

Table 2: *Comparison of different methods on Dataset II*

Method	*Within 100%*	*Within 120%*
RGB	68%	76%
Weighted	68%	76%
Transformed RGB	72%	80%
15 bin HSV Color Correlogram	71%	80%
Color Split	73%	80%
Weighted Variable Bin Size	73%	80%
Transf. RGB Color Correlogram	73%	93%

Figures 3 & 4 show the ranked results for two queries in Dataset I for the 15-bin HSV color correlogram feature. The top left retrieval with a rank of zero in each figure is the original query example. Figure 3 is the result of a query example of a yellow taxi. There were 14 vehicles labeled yellow in the ground truth data. Only 11 are returned in the first 15 retrievals. Figure 4 shows the results for a white car query in Dataset I. There are 21 white cars in the ground truth data. Twenty are retrieved in the first 20, but the last one is retrieved at rank 24.

Figure 3. Ranked results for query in top left. There were 14 vehicles labeled yellow in the ground truth. Only 11 are returned in the first 15.

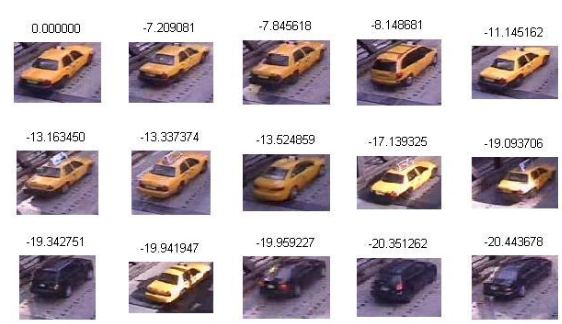
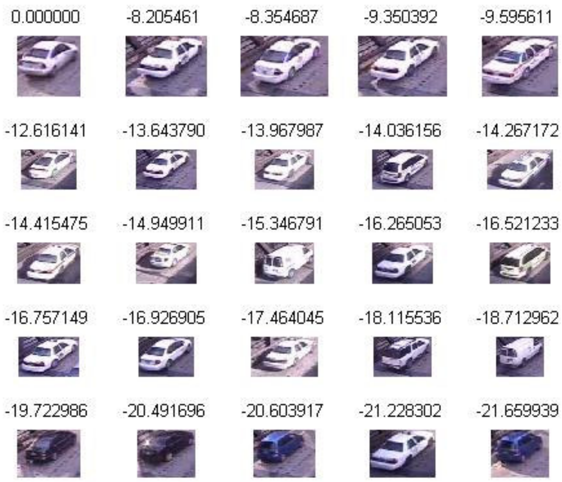


Figure 4. Ranked results for white car query. Several non-white cars are retrieved in the last row.



Figures 5 & 6 show the ranked results for queries in Dataset II using the color correlogram computed using the transformed RGB space with 512 bins. Figure 5 shows the results from Dataset II for retrieval of a red car. There are 4 red cars labeled in the ground truth. Notice the fourth red car does not appear until the last row (retrieval 23.)

Figure 6 shows the results for a white car. There are 62 vehicles labeled "light" in the ground truth data. The retrieval shows 74 returned cars (20% more than if all the light cars were returned in the first 62 places.) The "within 120%" score is 82% (or 61 cars.) The "within 100%" score is 74% (or 46 cars.)

Figure 5. Results of query for red car in Dataset II.

From Figure 6, it can be seen that lighting conditions play a significant role in challenging the system. We are experimenting with adding additional information from the background image of the road (acquired just prior to the event) in order to calibrate for the lighting conditions.

Figure 6. Ranked results for white car in dataset II which has 62 vehicles labeled "light" in the ground truth set.

5. Conclusions

We have designed a new functionality for surveillance systems based on real-time alerting and example-based specification of color. The user sets a tripwire to detect a vehicle crossing a certain virtual location. The user additionally selects an example of a vehicle of similar color to restrict the detection to vehicles in this color class We have shown that this functionality can provide accurate retrieval of vehicles of similar color distributions.

This is possible because (1) pose of the vehicles is limited by the tripwire specification, (2) lighting is less variable because of the pose, and (3) retrieval costs are minimal because testing is limited to the query-based example.

This framework has facilitated the evaluation of the efficiency and effectiveness of several color features including standard RGB and normalized RGB histograms, weighted, variable bin size, and color-split HSV histograms and color correlograms. Results show color correlogram to have the best performance for our datasets.

In the future, we plan to train a color classifier using a large dataset of vehicles from both single and varying poses with ground truth labels for vehicle color. We plan to use the features described in this paper either as global color features or as a small set of region features. This vehicle color classifier can then be used to label vehicles with their color for more extensive search and retrieval. We plan to develop a way to encode the histogram compactly to reduce storage and store it in a tree structure to facilitate efficient retrieval.

6. References

[1] R. Feris, et al., "Case Study: IBM Smart Surveillance System," Intelligent Video Surveillance: System and Technology, by Taylor & Francis Group, LLC, 2009.

[2] L.M. Brown, "Color Retrieval for Video Surveillance," Advanced Video and Signal Based Surveillance (AVSS), Albugquerque, NM, Sept 2009.

[3] M. V. Suhhamani and C.R. Venugopal, "Grouping and Indexing Color Features for Efficient Image Retrieval," *World Acadamy of Science, Engineering and Technology* 27, 2007.

[4] E. Zavesky, S-F Chang, "Columbia University's Semantic Video Search Engine 2008," *Proc. Of the 2008 Conf. on Image and Video Retrieval*, p545-546, Niagara Falls, Canada 2008.

[5] A. Natsev et al. "IBM Research TRECVID-2009 Video Retrieval System, TRECVID 2009.

[6] M. J. Swain, D. H. Ballard, "Color Indexing," *Int'l Journal of Computer Visio*n, Vol 7, No. 1, p11-32, 1991.

[7] K. E. A. van de Sande, T Gebers and C. G. M. Snoek, "Evaluation of Color Descriptors for Object and Scene Recognition," IEEE Pattern Analysis and Machine Intelligence (PAMI) (in press) 2010.

[8] N. Sebe and M. S. Lew, "Robust Color Indexing," Proc of the Seventh ACM Int'l Conf on Multimedia (Part 1), Orlando FL, p239-242, 1999.

[9] G. Qiu, J. Morris, and X. Fan, "Visual Guided Navigation for Image Retrieval," *Pattern Recognition*, Vol. 40, Issue 6, p1711-1721, June 2007.

[10] A. Mufit Ferman, A Murat Tekalp, R. Mehrotra, "Robust Color Histogram Descriptors for Video Segment Retrieval and Indentification," *IEEE Trans. On Image Processing*, Vol II, No. 5, May 2002.

[11] J. van de Weijer, T. Gevers, & A. Bagdanov, "Boosting Color Saliency in Image Feature Detection," *IEEE Trans. Pattern Analysis and Machine Intelligence* (PAMI), Vol 28, No. 1, p150-156, 2006.

[12] Z. Lei, et al., "A CBIR Method Based on Color-Spatial Feature," *Proc. IEEE Region 10 Annual International Conference 1999* (*TENCON'99*), Cheju, Korea. p166-169, 1999.

[13] D. Borhesani et al., "Color Features Performance Comparison for Image Retrieval," *Image Analysis and Processing* – ICIAP 2009, Lecture Notes in Computer Science, Springer Berlin/Heidelberg, p902-210, August 2009.

[14] Y. Deng et al., "An Efficient Color Representation for Image Retrieval," *IEEE Trans. on Image Processing*, Vol. 10, No. 1, p140-147, January 2001.

[15] G. Pass and R. Zabih, "Histogram refinement for content-based image retrieval," *IEEE Workshop on Applications of Computer Vision,* p96-102, 1996.

[16] J. Huang. et al., "Image Indexing Using Color Correlograms," *Proc. the 1997 Conf.on Computer Vision and Patter Recognition*, p762-768.

[17] T. Ojala, M. Rautiainen, E. Matinmikko, and M. Aittola. "Semantic image retrieval with hsv correlograms," *12th Scandinavian Conf. on Image Analysis*, pages 621–627, Norway, 2001.

[18] J. Hafner et al. "Efficient color histogram indexing for quadratic form distance functions." *IEEE Trans. On Pattern Analysis and Machine Intelligence,* Vol 17, No. 7, p729-736, 1995.