Comparative Performance Evaluation of Edge Histogram Descriptors and Color Structure Descriptors in Content based Image Retrieval

Pranoti Mane
Associate Professor and Head
Dept. of E&Tc. ,MESCOE
Pune,India

N. G. Bawane, Ph.D
Principal
S.B.Jain Institute of Technology,Management & Research,Nagpur

ABSTRACT
Content based image retrieval (CBIR) system is broadly used for searching and browsing images from a large database by extracting the visual content of the images. Image database is build by feature vectors corresponding to texture, color, shape and spatial features. The MPEG-7 standards provide standardized tools to describe and search audio and video contents. In this paper we apply the edge histogram descriptors (EHD) and color structure descriptors (CSD) standardized by MPEG-7 standard to different kind of images for content based image retrieval. We have used a database of 1000images provided by Wang.et.al. Experimental results are shown by calculating and plotting Precision and Recall performance parameters for three approaches which include the retrieval by EHD and CSD method individually as well as a combined similarity measures.

General Terms
Content based image retrieval.

Keywords
Content based image retrieval, Precision, Recall.

1. INTRODUCTION
Content based image retrieval is concerned with searching & browsing digital images from a database collection. CBIR is used to explore the research in many fields of image processing, multimedia, medical diagnosis and therapy, biomedical applications, agricultural development, astronomy and many more. To build an image database, feature vectors corresponding to texture, color, shape or spatial features are extracted and stored into a database[1][2][3][8]. When a query image is given, similar feature vector is computed and the next step involves matching these features to yield a result that is visually similar. The MPEG-7 standards provide standardized tools to describe and search video and audio contents. In this paper MPEG-7 color and texture descriptors are used independently as well as an integrated approach is used where both the descriptors are used together to retrieve the images from a large database containing various types of images.

Section 2 focuses on the MPEG-7 visual standards for color and texture features which are used in the experimentation. Color features are extracted by computing a color structure descriptor (CSD) as mentioned in section 3. The texture features are retrieved by EHD as discussed in section 4.

Experimentation and results of CSD, EHD and an integrated approach using both CSD as well as EHD by using standard database including 1000 different images is shown in section 5. Finally section 6 concludes the paper.

2. MPEG-7 VISUAL STANDARDS FOR COLOR AND TEXTURE FEATURE DESCRIPTION
The MPEG-7 is an ISO/IEC standard [9] developed by Moving Picture Experts Group (MPEG). It was formally named as “Multimedia Content description interface” and it provides standardized toolset to describe and search audio and video contents. There are more than 125 independent standards in MPEG group and a few of them are MPEG-1(Video and MP-3), MPEG-2(Digital T.V. set top box and DVD), MPEG-4(fixed and mobile web), MPEG-7(description and search of audio and video contents), etc. The rapid increase in audio and visual information has created a demand for representation that goes beyond the simple waveform or sample based, compression based (e.g.MPEG-1, 2) or object based (e.g. MPEG-4) representation. MPEG-7 focuses on the description of interpretation and content of the image and hence it plays a key role in content based image retrieval.

MPEG-7 visual description tools consist of basic structures and descriptors that cover basic visual features such as color, texture, shape, motion and domain specific features such as localization, identification of human faces and face recognition. Each category consists of elementary and sophisticated descriptors. This paper concentrates on the use of general visual descriptors such as color and texture mentioned in MPEG-7.

2.1. MPEG-7 color descriptors
Color is the most frequently used primitive and visual feature for the content based image retrieval. Color features are independent of image size, directionality and very robust to the changes in image background colors. There are seven color descriptors according to MPEG-7 standards [4][9] as mentioned below.

1. Color Space Descriptors: These descriptors define the selection of the color space to be used.
2. Color Quantization: These descriptors effectively specify partitioning (uniform quantization) of the given color space into discrete number of bins that provide a great flexibility for many applications.
3. Dominant Color Descriptors: Dominant color descriptors describe a small number of dominant color values and give their statistical aspects like distribution and variance. Thus it provides efficient, compact and perceptive representation of colors present in each region or image.
4. Scalable Color Descriptors (SCD): The SCD is a color Histogram defined in HSV (Hue, Saturation and Value) color space which uses a Haar Transform coefficient encoding.
5. Color Layout Descriptors (CLD): These descriptors effectively represent the spatial layout or spatial distribution

of dominant colors on a grid overlaid on a region of interest or an image. Discrete cosine transform is used to represent these descriptors.

6. Color structure descriptors (CSD): As its name implies, the CSD represents both color content information i.e. color histogram as well as the information about the structure of its content i.e. localized color distribution using structuring window. This method cements the color structure information into the CSD by considering all colors in a structuring window which slides over the image, instead of considering individual pixel separately [4][5]. To guarantee the interoperability, the CSD is bound to the double-coned HMMD color space which is quantized non-uniformly into 32, 64, 128 or 256 bins.

7. Group of frames / Group of Pictures color descriptors (GoF/GoP): It is nothing but the extended version of scalable color descriptor and used for color description of video segment or a collection of still images. Average median or intersection Histograms are calculated before the Haar transform is applied.

2.2. MPEG-7 texture descriptors

Texture is a broad-spectrum concept that recognizes more or less all aspects of nature. These natural objects can be described in texture images as the images of natural textured surfaces and synthetically generated visual patterns. There are three texture descriptors specified by MPEG-7 standards viz. texture browsing descriptors, homogeneous texture descriptors and edge histogram descriptors[9].

1. Texture Browsing Descriptors: This represents homogeneous texture for browsing type applications. These are very compact descriptors as they need only up to 12 bits to provide characterization of texture's regularity (2 Bits), directionality (3Bits X 2) and coarseness (2 Bits X 2). In this, the image is initially filtered with a bank of orientation and scale-tuned filters which use Gabor functions. Then two dominant texture orientations are found out from the filtered outputs. This texture browsing descriptor can be useful for high speed and more accurate image retrieval applications when combined with the homogeneous texture descriptors.

2. Homogeneous Texture Descriptors (HTD): The computations of these descriptors is on the basis of filtering the image using scale and orientation selective filters following which the mean and standard deviation of the filtered outputs is calculated in frequency domain. As HTD provides a quantitative characterization of texture for similarity based image to image matching, it is used for searching and browsing a large data base of texture images.

3. Edge Histogram Descriptors (EHD): Edge Histogram Descriptors are basically used for retrieval of natural images with non-uniform edge distribution or non-homogeneous texture. It considers the spatial distribution of 5 types of edges including 4 directional edges visibly vertical, horizontal, 45^0 diagonal, 135^0 diagonal and one non-directional or non-oriented edge. The original image is first divided into 16 sub images [8][4]. Then for each sub-image a local edge histogram which corresponds to 5 bins according to the orientation is computed. To refine the search or to make the retrieval more efficient here we have computed semi-global and global histograms along with local edge histogram.

3. CSD: COLOR STRUCTURE DESCRIPTORS

In this paper color structure descriptors are used for retrieving the images from data base containing different kinds of images of natural scenery, animals, human figures, buses, cars, flowers, etc. CSD expresses the image in both the color distribution in an image i.e. color histogram as well as local spatial structure of the color which explores certain image features which cannot be explored by color histogram. As mentioned in section I, the CSD expresses the local color structure in an image by using a structuring window and uses HMMD color space mentioned in MPEG-7 standards [5].

Table 1. Conversion of RGB to HMMD and HSV color space used in MPEG-7

Component	
Max (0 to1)	Max = max(R,G,B)
Min (0 to1)	Min= min(R,G,B)
Diff (0 to1)	Diff= Max-Min
Saturation (0 to1)	If (Max==Min) then Saturation=0; Else Saturation = (Max-Min)/Max

Hue (0^0 to 360^0)	If(Max==Min) then Hue is undefined(achromatic color);
	Otherwise:
	If(Max==R && G>B)
	Hue=60*(G-B)/(Max-Min)
	Else
	If (Max==R && G<B)
	Hue=360 + 60* (G-B)/(Max-Min)
	Else
	If (G==Max)
	Hue= 60* (2.0 + (B-R)/(Max-Min))
	Else Hue=60*(4.0 + (R-G)/(Max-Min))

Value (0 to1)	Value=max(R,G,B)

HMMD color space is defined in Three Dimensions i.e. Sum and diff axes and Hue angle. Its three dimensional quantization describes the partition of space into 3-D cells. As defined in MPEG-7 standards the four quantization partition the space into 256, 128, 64 & 32 cells respectively. Here we have partitioned the space into 128 cells. Each 3-D quantization is defined by 5 sub spaces of HMMD. The diff-axis which is defined on the interval 0 to 255 is divided into 5 sub-intervals Viz. [0, 6], [6, 20], [20, 60], [60, 110] & [110, 255]. These 1-D partitions define 5 different sub-spaces numbered from 0 to 4 respectively where sum & Hue includes all the values in respective ranges but diff is restricted to one of the 5 sub-intervals given in Table 2.

Table 2. HMMD Quantization

Subspaces	HMMD Quantization---Number of Cells into the Subspaces							
	256		128		64		32	
	Hue	**Sum**	**Hue**	**Sum**	**Hue**	**Sum**	**Hue**	**Sum**
0	1	32	1	16	1	8	1	8
1	4	8	4	4	4	4	4	4
2	16	4	8	4	4	4		
3	16	4	8	4	8	2	4	1
4	16	4	8	4	8	1	4	1

As we have partitioned the HMMD space into 128 cells, according to the Table 2, subspace 4 is to be partitioned by cutting Hue into 8 uniform intervals and Sum into 4 uniform intervals. As mentioned in the above table the remaining four subspaces are to be partitioned to get an overall non-uniform quantization of HMMD space [5].

The color structure descriptor consists of one dimensional array of 8 bits non-uniformly quantized values.

$\overline{CSD} = \overline{h_s}\,(m)$, where $m = \{1,2,3,\ldots,M\}$ and M is selected from a set of{256,128,64,32} and s is the scale of square structuring element. We have selected M=128 bins. The selection of spatial extent of structuring element is done according to the rule as given below [4].

$p = \text{maximum}\{ 0, \text{round}(0.5*\log_2 WH - 8) \}$

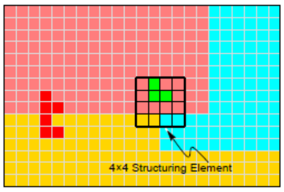
$k = 2^p$; $E = 8*k$

Where W is the width of the image
H is the height of the image
K is sub sampling factor
E X E is the spatial extent (size of) structuring element.

As the image size is selected as 256 X 256, according to above criterion, p= 0; k=1 and hence E=8. Therefore size of structuring element (SE) or structuring window is selected as 8 X 8.

CSD Extraction: The 8 X 8 SE slides over the image and counts the number of times a particular color is contained within itself. A 128 bin color structure histogram is extracted from an image represented in 128 cell quantized color space. If the image is in another color space, then it is first converted into HMMD color space and re-quantized prior to extraction.

Computation of a CS Histogram: The procedure of computation of CS histogram is illustrated by taking an example of an image which contains five different colors as shown in Fig. 1. The size of structuring element is chosen as 4 X 4. The table beside the figure shows an 8-bin CS Histogram, *hs*(*m*), whose bins are associated with 8 quantized colors, c_m (m varies from 1 to 8) by which the image is represented.



Color	Bin Value
c_1	$h(1)$ +1
c_2	$h(2)$
c_3	$h(3)$
c_4	$h(4)$ +1
c_5	$h(5)$
c_6	$h(6)$ +1
c_7	$h(7)$ +1
c_8	$h(8)$

Fig. 1. Computation of Color structure Histogram [5]

The SE slides over the image and scans the image. At each location the CS Histogram is updated when a particular color of image is present within the element. At that location, the increase in $h(m)$ depends on presence of color c_m within the structuring element rather than the quantity of c_m is enclosed. In this way the final value of $h(m)$ is decided (up to normalization) by the number of locations at which the structuring window contains c_m .[5]

4. EHD: EDGE HISTOGRAM DESCRIPTORS

Edge histogram descriptors defined by MPEG-7 standards are mainly used for retrieval of non-uniform edge distribution or non-homogeneous texture images. These descriptors describe the distribution of edges in an image with the help of a histogram based on local edge distribution [6][7][7]. This local edge histogram is computed by dividing the original image into 16 non-overlapping blocks of equal size. The edge information is extracted from each block and represented in terms of local histogram containing five bins for each block. Each of these five bins depicts the information either of the edges in Vertical edge, horizontal edge, 45 degree diagonal edge, 135 degree diagonal edge and non-directional edge. The histogram for each block or sub-image specifies the relative frequency of occurrence of the five types of edges and the directionality of the brightness changes in the image in that particular block. The total number of bins computed for the whole image is 16 X 5 = 80 (16 sub-images and 5-bins) and the one-dimensional array of 80 bins of EHD is constructed. Each bin specifies its own location and edge type. The procedure of extraction of edge features from an image is given as follows[7].The original image is first divided into 16 non-overlapping image blocks of equal size. Each image block is then split into four sub-blocks and assigned the labels from 0 to 3. Average gray level for each block is then calculated as $a_k(i,j)$, where k is the location of the sub-block(k=0 to 3) and (i,j) specifies the location of image block(i=0 to 3 and j=0 to 3). For example $a_0(3,3)$ identifies the average gray level of the zero numbered sub-block of the image block which is located at the right most- bottom corner of the original image. For each direction, filter co-efficient is represented as $f_v(k)$, $f_h(k)$, $f_{d-45}(k)$, $f_{d-135}(k)$ and $f_{ND}(k)$ for vertical,horizontal,45^0 diagonal, 135^0 diagonal and non-directional edges respectively. Each of the image blocks is then classified into one of the five mentioned edge categories or as a non directional edge block.

$f_v(k) = \{1 -1\ 1\ -1\}$, $f_h(k) = \{1\ 1\ -1\ -1\}$, $f_{d-45}(k) = \{\sqrt{2}\ 0\ 0\ \sqrt{2}\}$, $f_{d-135}(k) = \{0\ \sqrt{2}\ \ \sqrt{2}\ \ 0\}$ and $f_{ND}(k) = \{2\ -2\ 2\ -2\}$

Now the respective edge strengths are computed by using the following equations [7] .

$$m_v(i,j) = \left| \sum_{k=0}^{3} a_k(i,j) \times f_v(k) \right| \quad (1)$$

$$m_h(i,j) = \left| \sum_{k=0}^{3} a_k(i,j) \times f_h(k) \right| \quad (2)$$

$$m_{d-45}(i,j) = \left| \sum_{k=0}^{3} a_k(i,j) \times f_{d-45}(k) \right| \quad (3)$$

$$m_{d-135}(i,j) = \left| \sum_{k=0}^{3} a_k(i,j) \times f_{d-135}(k) \right| \quad (4)$$

If the maximum out of the edge strengths(equation (5)) is greater than a threshold then the image block is considered to have this edge in it otherwise it contains no edge[7]. Here the threshold is selected as 0.1.

$$\max\{m_v(i,j), m_h(i,j), m_{d-45}(i,j), m_{d-135}(i,j), m_{nd}(i,j)\} > T_{edge}$$

…. (5)

The 80 bins of local histogram are represented by bincounter[i] where i varies from 0 to 79. For making the image retrieval more efficient, some global and neighbor (Semi-global) distributions [7][7] are also used and the distance between the database image and query image is computed as per the given equation.

$$\textbf{Dist (D, Q)} = \textbf{L*} \sum_{i=0}^{79} |L\_D[i] - L\_Q[i]| + \textbf{G*} \sum_{i=0}^{4} |G\_D[i] - G\_Q[i]| + \textbf{N*} \sum_{i=0}^{64} |SG\_D[i] - SG\_Q[i]|$$

Where L=Local weight (Assigned as 1).
G= Global weight (Assigned as 5 since the number of bins of global histogram is relatively smaller than that of local and semi-global).
N= Semi-global weight (Assigned as 1).
L_D[i] & L_Q[i] are the reconstructed values of bincounter[i] of database and query image respectively.
G_D[i] & G_Q[i] are the normalized histogram bin values for the global edge histogram of database and query image respectively.

SG_D[i] & SG_Q[i] represent the normalized histogram bin values for the semi-global or neighboring groups- edge histogram of database and query image respectively.

Semi-global edge histogram is obtained from image blocks by considering four rows, four columns and five groups of 2X2 blocks as mentioned below.

4 rows : Contain subset of image blocks (1,2,3,4),(5,6,7,8),(9,10,11,12) and (13,14,15,16) .

4 columns: Contain subset of image blocks (1,5,9,3),(2,6,10,14),(3,7,11,15) and (4,8,12,16) .

5 groups of 2X2 blocks: contains subset of image blocks (1,2,5,6),(3,4,7,8),(9,10,13,14),(11,12,15,16) and (6,7,10,11) .

The main edge histogram is computed by using the local, global and semi-global groups.

5. EXPERIMENTATION AND RESULTS OF CSD, EHD AND AN INTEGRATED APPROACH

Standard database of 1000 images [Wang et.al] is used and the primitive features i.e. Color and texture are extracted by using Edge Histogram Descriptor and Color Structure Descriptor. Also the same features are extracted from the query image and stored. The distance between the query image and all database images is computed by the above mentioned procedures and arranged in ascending order in an array. The evaluation was done by using CSD and EHD separately. Top 20 similar images are displayed and evaluation is carried out. An integrated approach is also used which makes use of EHD in conjunction with CSD for the same query images and the images are displayed. The performance parameters Precision and Recall are evaluated for all the three approaches, i.e. only color feature extraction (CSD), only Texture feature extraction (EHD) and an integrated approach (color and texture feature extraction). Results are shown in figure 2 and figure 3.The performance parameters, Precision and Recall are defined as follows. [10][11][12]

Precision = No. of Relevant images retrieved / Total No. of images retrieved

Recall = No. of Relevant images retrieved / Total No. of relevant images in the database.

6. CONCLUSION

In this work, we have presented and evaluated the techniques for computing color and texture similarities using Color Structure Descriptors and Edge Histogram descriptors. The retrieval effectiveness is evaluated in the experiment of retrieving 10 different images from a set of 1000 images of a standard database. The results are different for different types of images. Also the results are depending on the method used. Whenever there is a distinct object on a background like all buses or human figures then EHD method is outperformed. For the image containing principal colors, CSD method shows good results. If there are principal colors and distinct objects in an image then image retrieval by combined approach gives good results. All the methods give similar precision and retrieve similar images when the database contains the images having similar texture and color of the object and similar background as shown by dinosaur image retrieval. In case of a query image which contains many objects like the horse, pony, fencing and green field, the results are better by CSD and both (integrated) method individually. It concludes that the integrated approach of color and texture works well in multiple object images. For some images like food plate all the methods are showing very poor results though various food plate images are present in the database.

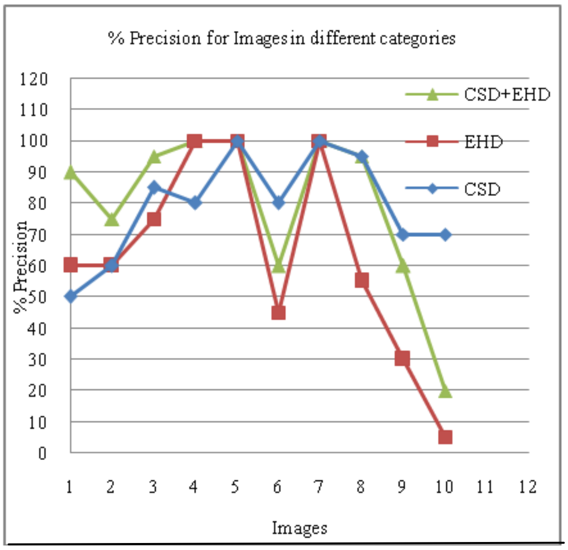


Figure 2:%Precision for images in different categories

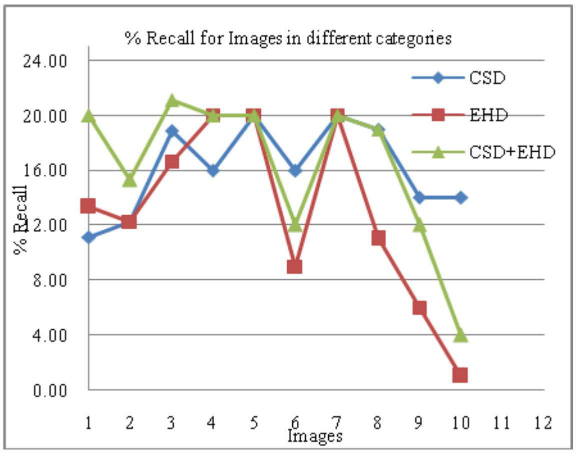


Figure 3: %Recall for images in different categories

These efforts in image retrieval applications have focused on only low level or primitive feature extraction such as color and texture. However without integrating these techniques with relevance feedback [13], the current CBIR methods have limited capacity to satisfactorily retrieve more semantically meaningful images.

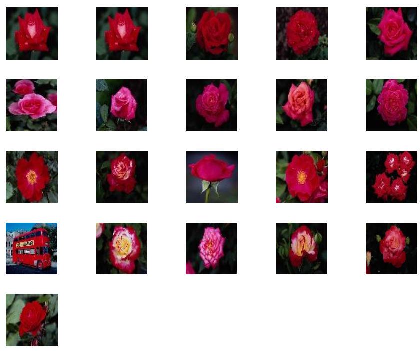


Figure 4: Top 20 retrieved images of flower by CSD method. First image is the Query image. Precision=19/20, Recall=19/100

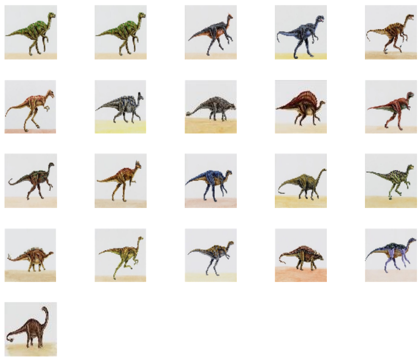


Figure 5: Top 20 retrieved images of dinosaur by EHD method. First image is the Query image. Precision=20/20, Recall=20/100

7. REFERENCES

[1] Ying Liu,Dengsheng Zhang,Guojun Lu and Wie- Ying Ma,“A survey of content-based image retrieval with high-level semantics”, Pattern Recognition,vol.40,issue1,January 2007,pp262-282.

[2] M.B .Kokare, B.N. Chatterji and P.K. Biswas, “A Survey On Current Content Based Image Retrieval Methods”, IETE Journal of Research, 2002,pp. 261-271.

[3] G.Rafiee, S.S.Dlay, and W.L.Woo, “A Review of Content-Based Image Retrieval”, CSSNDSP 2010,pp.775-779.

[4] B. S. Manjunath, Member, Jens-Rainer Ohm, Member, Vinod V. Vasudevan, and Akio Yamada, “Color and Texture Descriptors”, IEEE Transactions on Circuits And Systems For Video Technology, vol. 11, No. 6, June 2001, pp.703-715.

[5] J.R. ohmer, Heon Jun Kim, S. Krishnamachari ,B. S. Manjunath , Akio Yamada , “The MPEG-7 Color Descriptors”.

[6]D.K.Park,Y.S.Jeon,C.S.Won, “Efficient Use of Local Edge Histogram Descriptor”, Proceeding, ACM workshops on Multimedia, November 2000, pp.51-54.

[7] C.S.Won,D.K.Park,S,J.Park, “Efficient Use of MPEG-7 Edge Histogram Descriptor”, ETRI Journal, vol. 24, Number 1, February 2002,pp.23-30

[8] Thomas Sikora, “The MPEG-7 Visual Standard for Content Description—An Overview”, IEEE Transactions on Circuits and Systems for Video Technology, vol. 11, No. 6, June 2001, pp.696-702

[9] Palma de Mallorca, October 2004, “MPEG-7 Overview”, ISO/IECJTC1/SC29/WG11N6828.

[10] Vasileios Mezaris1,2, Ioannis Kompatsiaris2, and Michael G. Strintzis, “An Ontology based approach to object based Image retrieval” ,ICIP 2003.

[11] Samuel Barretta, Ran Changb, and Xiaojun Qib, “A Fuzzy based learning approach to CBIR”,ICME2009, IEEE, pp 838-841.

[12] Pierre Blanchart and Mihai Datcu, “ A Semi-Supervised Algorithm for Auto-Annotation and unknown Structures Discovery in Satellite Image Databases”, IEEE journal of selected topics in applied earth observations and remote sensing, vol. 3, No. 4, December 2010,pp698-717.

[13] Yong Rui, Thomas S. Huang, Michael Ortega and Sharad Mehrotra “Relevance Feedback: A Power Tool for Interactive Content-Based Image Retrieval”, IEEE Transactions on Circuits and Video Technology, vol. 8,issue 5, pp.644-655.

[14] Mohammed Lamine Kherfi and Djemel Ziou, “Relevance Feedback for CBIR: A New Approach Based on Probabilistic Feature Weighting with Positive and Negative Examples”, IEEE Transactions on Image Processing, vol. 15, No. 4, April 2006, pp.1017-1030.