THE ANGULAR ORIENTATION PARTITION EDGE DESCRIPTOR

António M. G. Pinheiro
Unidade de Detecção Remota, Universidade da Beira Interior
6200 Covilhã - Portugal

ABSTRACT
Edges are one of the most important image visual features. They are highly related with shapes and can also be representative of the image textures. Edge orientations histograms are usually very reliable descriptors suitable for image analysis, search and retrieval. In this work edges detected with Canny algorithm are described by their angular orientations. The resulting descriptor is resilient to image rotation and image translation. It is also resilient to noise.

An example of automatic image semantic annotation using this description method is reported using a database with 738 images. The K Nearest Neighbor is used as classifier and the Manhattan distance is used for image similarity computation. The annotation that results with this description method is compared with the provided with other well known descriptors. These examples show that a reliable high level automatic description based in the semantic content can be extracted.

Index Terms— Image classification, Image database

1. INTRODUCTION
Multimedia information retrieval is one of the most important challenges of multimedia technology. Recently a strong focus has been put on the research of techniques that “bridge the semantic gap” between the extraction of low level audio-visual features and the production of high level descriptors representative of the multimedia information semantic content [5]. The first steps of research on multimedia information retrieval have been dominated by the development of retrieving methods based on query by example. However, soon was concluded that there is a need of an image analysis based in its content. Although the huge number of description techniques recently researched and published, only few represent reliable solutions suitable for image semantic annotation. Some are relevant because they give a very positive contribution to reduce the “semantic gap”. Oliva and Torralba [9] proposed a Spatial Envelope to model the shape of the scene, that results in an effective image semantic annotation. Lowe [6] proposes the SIFT descriptor invariant to image transformations based on keypoint detection. Those keypoints orientation and gradient is computed to form a 3D descriptor, used for image comparison. Several local descriptors, as SIFT, other keypoint descriptors based on the Harris detector, or also a SIFT variation, are studied and compared in [8].

Edge orientations histograms always represented a middle term between reliability and description dimension. They are often used in practical applications because they provide very efficient and simple solutions. As an example, the Edge Histogram Descriptor (EHD) [7] of the MPEG-7 standard, is one of the most popular descriptors, and probably the most used descriptor of the standard. One of the most criticized limitations of edge description is that it is not resilient when image rotations are present.

In this paper, the Angular Orientation Partition Edge Descriptor (AOP) suitable for image semantic annotation, and resilient to image rotation and translation, is described. It is based on the Angular Radial Partition Descriptor (ARP) [2]. When used for image semantic annotation based in image classification, the ARP descriptor results in a very poor performance. However, adding new considerations, like separating edge pixels into different edge orientations and effectively studying all the parameterization involved, a very reliable descriptor that might be used for semantic annotation and eventually several other computer vision applications, is designed.

2. THE ORIGINAL ARP DESCRIPTOR
The ARP descriptor [2, 3] was originally designed as an edge based descriptor resilient to rotation. Edges are detected with Canny algorithm [1]. The edge image is divided into circular sectors of the surrounding circle (see figure 1(a)). The algorithm defines \( N_r \) radial divisions and \( N_a \) angular divisions resulting in \( N_r \times N_a \) sectors. The number of edge pixels in each sector \((n_r, n_a)\) is computed, \( f(n_r, n_a) \), with \( n_r = 0, 1, ..., N_r - 1 \) and \( n_a = 0, 1, ..., N_a - 1 \). This image feature, \( f(n_r, n_a) \), is circularly shifted for an image rotation of \( \alpha = 2\pi n/N_a \) with \( n \) integer.

Computing the 1-D Discrete Fourier Transform relatively to the dimension \( n_a \) of the feature \( f(n_r, n_a) \) results in \( F(n_r, \omega_a) \). For the rotated image results

\[
F_\alpha(n_r, \omega_a) = F(n_r, \omega_a) e^{-j2\pi n_\omega a/N_a}
\]  

(1)

Then, the absolute values of \(|F(n_r, \omega_a)| = |F_\alpha(n_r, \omega_a)|\) are
equal. So, the feature $|F(n_r, \omega_0)|$ is chosen as the image descriptor, and it is invariant to image rotations multiples of $2\pi/N_\omega$. [2, 3] claims a very good performance for image similarity computation resilient to image rotation. However, the descriptor exhibits a real bad performance when used for image semantic classification.

3. THE DEVELOPED AOP EDGE DESCRIPTOR

New characteristics are added to the original ARP descriptor to increase the image comparison reliability. The new attributes are: 1) Edge orientation (the direction is also considered) information added to the descriptor; 2) Modified gradient image mass center computation for more reliable image comparison; 3) Moreover, no radial division is done. The first characteristic separates kinds of edge points, considering the angular orientation. The feature $f(n_o, n_a)$ counts the number of edge pixels with orientation $n_o$ in the angular division $n_a$, $n_o = 0, 1, ..., N_o - 1$ represents each of the defined pixel edge angular orientations (see figure 1(b)), where $N_o$ is the number of orientations considered. The orientation is given by the image gradient angular orientation in the edge pixel, and is computed rotating the cartesian axes (the horizontal axis is rotated to the radius that includes the edge point). The angular orientation is then computed by:

$$
\Delta_a = \Delta_x \cos(\theta) - \Delta_y \sin(\theta) \quad \text{Parallel to radius}
$$

$$
\Delta_n = \Delta_x \sin(\theta) + \Delta_y \cos(\theta) \quad \text{Normal to radius}
$$

considering $\Delta_x$ and $\Delta_y$ the derivative approximations in the horizontal and vertical directions, and $\theta$ the angular orientation (between the line joining the edge point and the image division central point, and the horizontal positive direction). An interval $n_o$ (each interval has $2\pi/N_o$ radians) that includes the angular orientation angle given by $\tan^{-1}(\Delta_n/\Delta_r)$ is computed. As an example, if the number of orientations is $N_o = 4$, then the orientations $n_o = 0, 1, 2, 3$ are defined. Each edge pixel will have respectively one of closer angular orientations $0, \pi/4, \pi/2, 3\pi/4$. The final descriptor will be given by $|F(n_o, \omega_0)|$ that results from taking the absolute value of the 1-D discrete Fourier transform of $f(n_o, n_a)$ computed relatively to the angular dimension $n_a$. It is important to notice that as the orientations are computed relatively to the angular orientation, they are independent of $n_a$.

The modified gradient image mass center is also computed to define the center of the image angular division. The gradient edge image center results from computing the mass center of the image considering only the points where edges exist. It is given by:

$$
y_{mc} = \frac{\sum_{y,x} |\hat{g}|.E.y}{\sum_{y,x} |\hat{g}|.E}, \quad x_{mc} = \frac{\sum_{y,x} |\hat{g}|.E.x}{\sum_{y,x} |\hat{g}|.E}
$$

where $\hat{g}$ is the image gradient in point $(y, x)$, $E$ is 1 if $(y, x)$ is an edge point or 0 if it is not, and finally $\sum_{y,x}$ represents a summation all over the image. In figure 1(b), the circles are not centered in the image center because they have been moved to the modified gradient image mass center.

This way is expected to add resilience to image translation. Considering this factor it is also expected that a resilient semantic annotation for large image databases might result, using this descriptor.

The descriptor normalization, $|F_N(n_o, \omega_0)|$, is made over the absolute value of the 1-D discrete Fourier, transform $|F(n_o, \omega_0)|$, computed for each of the defined angular orientations. Those bins representative of each of those sub-sets are divided by the maximum bin of each sub-set (with the same angular orientation):

$$
|F_N(n_o, \omega_0)| = |F(n_o, \omega_0)|/\max \{|F(n_o, \omega_0)|\}
$$

for each orientation $n_o = 0, 1, ..., N_o - 1$. Hence, the descriptors is not normalized as a whole but by descriptors sub sets. However, the comparison between two different descriptors is made as a whole. In this work the Manhattan distance was used to compute the similarity between two image descriptors.

4. RESULTS

The description method was tested using a database with 738 kframes images of the TRECVID 2008 development database. The small set is used because it allows an easy testing and a private reliable annotation. An example of the detection of the image content “images with at least one building” is reported. The influence of the different AOP parameters in the classification performance, is studied using this example.

Images will be classified using the $k$ nearest neighbor algorithm ($k$NN) [4] with $k = 7$. The different points representative of the Precision versus Recall curves are obtained for different confidence intervals. The confidence interval is computed using the relation between $k_{pos}/k$, where $k_{pos}$ is the number of positive matches in $k$ that result from the application of the $k$NN algorithm.

3. THE DEVELOPED AOP EDGE DESCRIPTOR

New characteristics are added to the original ARP descriptor to increase the image comparison reliability. The new attributes are: 1) Edge orientation (the direction is also considered) information added to the descriptor; 2) Modified gradient image mass center computation for more reliable image comparison; 3) Moreover, no radial division is done. The first characteristic separates kinds of edge points, considering the angular orientation. The feature $f(n_o, n_a)$ counts the number of edge pixels with orientation $n_o$ in the angular division $n_a$, $n_o = 0, 1, ..., N_o - 1$ represents each of the defined pixel edge angular orientations (see figure 1(b)), where $N_o$ is the number of orientations considered. The orientation is given by the image gradient angular orientation in the edge pixel, and is computed rotating the cartesian axes (the horizontal axis is rotated to the radius that includes the edge point). The angular orientation is then computed by:

$$
\Delta_a = \Delta_x \cos(\theta) - \Delta_y \sin(\theta) \quad \text{Parallel to radius}
$$

$$
\Delta_n = \Delta_x \sin(\theta) + \Delta_y \cos(\theta) \quad \text{Normal to radius}
$$

considering $\Delta_x$ and $\Delta_y$ the derivative approximations in the horizontal and vertical directions, and $\theta$ the angular orientation (between the line joining the edge point and the image division central point, and the horizontal positive direction). An interval $n_o$ (each interval has $2\pi/N_o$ radians) that includes the angular orientation angle given by $\tan^{-1}(\Delta_n/\Delta_r)$ is computed. As an example, if the number of orientations is $N_o = 4$, then the orientations $n_o = 0, 1, 2, 3$ are defined. Each edge pixel will have respectively one of closer angular orientations $0, \pi/4, \pi/2, 3\pi/4$. The final descriptor will be given by $|F(n_o, \omega_0)|$ that results from taking the absolute value of the 1-D discrete Fourier transform of $f(n_o, n_a)$ computed relatively to the angular dimension $n_a$. It is important to notice that as the orientations are computed relatively to the angular orientation, they are independent of $n_a$.

The modified gradient image mass center is also computed to define the center of the image angular division. The gradient edge image center results from computing the mass center of the image considering only the points where edges exist. It is given by:

$$
y_{mc} = \frac{\sum_{y,x} |\hat{g}|.E.y}{\sum_{y,x} |\hat{g}|.E}, \quad x_{mc} = \frac{\sum_{y,x} |\hat{g}|.E.x}{\sum_{y,x} |\hat{g}|.E}
$$

where $\hat{g}$ is the image gradient in point $(y, x)$, $E$ is 1 if $(y, x)$ is an edge point or 0 if it is not, and finally $\sum_{y,x}$ represents a summation all over the image. In figure 1(b), the circles are not centered in the image center because they have been moved to the modified gradient image mass center.

This way is expected to add resilience to image translation. Considering this factor it is also expected that a resilient semantic annotation for large image databases might result, using this descriptor.

The descriptor normalization, $|F_N(n_o, \omega_0)|$, is made over the absolute value of the 1-D discrete Fourier, transform $|F(n_o, \omega_0)|$, computed for each of the defined angular orientations. Those bins representative of each of those sub-sets are divided by the maximum bin of each sub-set (with the same angular orientation):

$$
|F_N(n_o, \omega_0)| = |F(n_o, \omega_0)|/\max \{|F(n_o, \omega_0)|\}
$$

for each orientation $n_o = 0, 1, ..., N_o - 1$. Hence, the descriptors is not normalized as a whole but by descriptors sub sets. However, the comparison between two different descriptors is made as a whole. In this work the Manhattan distance was used to compute the similarity between two image descriptors.

4. RESULTS

The description method was tested using a database with 738 kframes images of the TRECVID 2008 development database. The small set is used because it allows an easy testing and a private reliable annotation. An example of the detection of the image content “images with at least one building” is reported. The influence of the different AOP parameters in the classification performance, is studied using this example.

Images will be classified using the $k$ nearest neighbor algorithm ($k$NN) [4] with $k = 7$. The different points representative of the Precision versus Recall curves are obtained for different confidence intervals. The confidence interval is computed using the relation between $k_{pos}/k$, where $k_{pos}$ is the number of positive matches in $k$ that result from the application of the $k$NN algorithm.

"http://www-alpir.nist.gov/projects/trecvid/"
The training set is a sub-set of the 738 kframes selected. They are not used in the final classification, otherwise results would be polarized by the training samples. In this case 50 negative and 25 positive training samples are selected. These images, that were selected in a previous work, are not chosen specially to be used together with this descriptor.

4.1. Performance analysis of the AOP

In this section the performance of the descriptor when used for semantic annotation of “images with at least one building” is analysed through the variation of the different descriptor parameters. Precision-Recall plots are used to allow an easy comparison.

i) Number of Orientations - Figure 2(a): Very similar performances are found for the cases of \( N_o = 8, 16 \) or 32 and it is degraded for \( N_o = 4 \). Nevertheless, a choice of \( N_o = 8 \) is the most robust and it is used in general. Moreover 16 or 32 orientations will result in a multiplication of the number of bins by 2 or 4 respectively, which is an obvious disadvantage.

ii) Orientation versus Direction and Modified Gradient Image Mass center versus Image center - Figure 2(b): Using Orientation instead of Direction results in a small improvement of the classification performance and also has the advantage of resulting in half length descriptors (8 orientations result in 16 directions). Using the gradient image mass center as the center of the angular description also results in slight better performance than using the image geometric center.

iii) The influence of the scale of the Canny algorithm gaussian filtering - Figure 2(c): The classification performance tends to degrade with larger scale \( t \) values (\( \sigma = \sqrt{2t} \)). However, scales \( t = 1 \) or 2 result in very similar performance. \( t = 1 \) was selected because it results in more efficient gaussian filtering computation.

iv) The influence of the number of image divisions - Figure 2(d): The performance with the number of angular division \( N_a = 8, 12, 16 \) and 24 is shown. The performance for a radial division like the suggested in the original ARP descriptor is also shown (\( N_r = 2 \)). Image radial divisions always resulted in performance degradation. The parameter \( N_a \) does not have a strong influence in the classification performance. A choice of \( N_a = 16 \) was made, because it results in slight better performance, although it duplicates the number of the descriptor bins when compared with \( N_a = 8 \).

v) The influence of the Canny algorithm hysteresis thresholding - Figure 2(e): Better classification performance results when no hysteresis thresholding is applied to the edge images. This fact is an advantage, because avoiding hysteresis thresholding improves the descriptor computation efficiency.

4.2. Comparison with other descriptors

A new descriptor is only significant if it provides a performance at least comparable with other known descriptors in a set of applications. Figure 2(f) shows a comparison with the very well known MPEG-7 EHD and SIFT. The original ARP (with a real poor performance) and the Edge Pixel Direction Histogram (EPDH) [10] Precision Recall curves are also shown. The experiments were made using Euclidean distance computation. In the case of SIFT, the implementation provided by Lowe in his personal home page was used. The number of matching points is used as similarity measure, for the \( k \)NN. Apart the EPDH [10], all the descriptors exhibit a poorer performance when compared with the developed AOP.

The AOP largely overcomes the performance of the original ARP. In fact, the ARP is only good for recognition of images that suffer a rotation centered in its geometric center. Any other application is not effective. SIFT is very powerful to find the same objects in different images. However, the image set selected from the TRECVID database, does not have any scene repetition, and the training sets have been selected only as meaningful examples. In most of the cases, SIFT does not find any similarity between the image under classification and the training images. The EHD of MPEG-7 also results in poorer classification performance. Finally, the EPDH has a performance comparable with the AOP descriptor. However, like the EHD, this descriptor is not resilient to image translation and rotation. The EPDH were computed using a image subdivision for the two edge images of \( 4 \times 4 \), making the two scales equal \( t = 2 \), which improves the performance of [10].

4.3. Resilience to rotation, translation and noise

To test the resilience of this method to rotation and translation, all database images were rotated with different random values of the rotation angle (around the image center) followed by a random translation between ± 10 pixels in vertical and horizontal orientations. Descriptors were computed for those images randomly transformed. Those descriptors have been compared with the original ones. 99% of the original images (731 out of 738) had as similar images the transformed image. The remaining 7 (1%) images have resulted as the second closer. In addition, after rotating and translating the images was added ±1% of noise (from 0 to 255, the pixels where randomly added with a gray value of ± 0, 1, 2 or 3). The same test resulted again in exactly the same results. These results (other similar tests have similar results) allow to conclude that this descriptor is very resilient to image rotation, translation and noise.

5. CONCLUSION

A new image descriptor resilient to image rotation and translation has been described in this paper. This method is also very resilient to image noise.

The descriptor performance depends of a large set of parameters. However, the classifications performance exhibits a very stable behavior when parameters are inside defined reliable intervals. The represented plots also show that the
developed descriptor performance always overcome the performance that results with other well known descriptors. The AOP descriptor is specially suitable for large databases annotation because it has a very efficient computation and allows a real fast comparison. Considering efficiency and classification performance the choice of the parameters used for the descriptor computation is: \( N_a = 16 \) or 8, \( N_d = 8 \), no radial division, no hysteresis thresholding and \( t = 1 \) or 2. It is also used angular orientation instead of direction.

As future research, the ability of the descriptor to describe image regions with arbitrary shapes will be considered.

6. REFERENCES