Objects recognition using SIFT and fuzzy similarity measure

Karddouchi, M.; Belacel, Nabil

Publisher’s version / Version de l'éditeur:

The 20th Annual IS&T/SPIE Symposium on Electronic Imaging [Proceedings], 2008

National Research Council Canada
Conseil national de recherches Canada

Institute for Information Technology
Institut de technologie de l'information

NRC · CNRC

*Objects recognition using SIFT and fuzzy similarity measure **

Karddouchi, M., Belacel, N.
January 2008

* published at The 20th Annual IS&T/SPIE Symposium on Electronic Imaging. San Jose, California, USA. January 27-31, 2008. NRC 49914.

Canada

Objects recognition using SIFT and fuzzy similarity measure

Mustapha Kardouchi[a] and Nabil Belacel[b]

[a]Computer Science Department, Université de Moncton
Moncton, NB, E1A 3E9, Canada
mustapha.kardouchi@umoncton.ca

[b]NRC, Institute for Information Technology, 55 Crowley Farm Road,
Suite 1100, Moncton, NB, E1A 7R1, Canada,
Nabil.Belacel@nrc-cnrc.gc.ca

ABSTRACT

Multimedia database has been an extremely active area of research over the last 20 years. This research aims to develop techniques for searching and recognizing multimedia documents based on their contents. For objects recognition, many works proposed different techniques to extract visual contents such as color, shape, texture etc. By comparing these visual contents, we can determine whether or not the image data contains some specific object. Since the Euclidean distance has always been employed to compare these visual contents so far, using other approaches for the comparison is an interesting research way that still needs to be explored. In this paper, we propose fuzzy similarity measures as alternatives for the Euclidean distance. Visual contents to be compared are based on the Scale Invariant Feature Transform (SIFT). This approach has been applied to coil databases. Our experimentation shows that this method is more realistic than objects recognition method obtained by classical distances.

Keywords: SIFT descriptors, fuzzy similarity, object recognition, fuzzy triangular number

1. INTRODUCTION

Content Based Image Retrieval (CBIR) based on visual content is a fundamental problem in computer vision. The basic idea of CBIR is to compactly describe an image by using combinations of primitive features such as color, texture and features points. These images descriptors are used with a similarity measure to retrieve images that are alike. The techniques used in CBIR have successfully been applied to objects recognition. Local approaches have demonstrated large success in a several class of applications, like image retrieval [1][2][3] and objects recognition [4][5], etc. They are commonly employed because they can be computed efficiently, are resistant to partial occlusion, and are relatively insensitive to changes in viewpoint [6][7].

There are several similarity measures that are proposed to measure similarity between local descriptors [8][9]. So far, the Euclidean distance has often been used to compare feature vectors of these types. In this paper, however, we identify the feature vectors with fuzzy triangular numbers, such that fuzzy similarity measures can determine the similarity between them. The fuzzy similarity measure is used as an alternative to the traditional Euclidian distance to recognize an object from a library of single objects. The main contribution of this paper is the introduction of new algorithm for image object recognition, which is based on the Scale Invariant Feature Transform (SIFT) descriptors [10] and fuzzy similarity measure [11]. SIFT transforms image data into scale invariant coordinates relative to local features. The features are invariant to image scale and rotation, change in 3D viewpoint, addition of noise, and change in illumination. For image recognition, SIFT features are first extracted from a set of reference images and stored in database. A new or a target image is recognized by comparing feature vectors from the new image to this previous database and finding the fuzzy similarity degree between them.

The remainder of this paper is organized as follows. Section 2 reviews the SIFT descriptors. Section 3 presents our approach based on Scale Invariant Feature Transform (SIFT) descriptors and fuzzy similarity measure. Section 4 details the fuzzy similarity measures used in this work. Section 5 provides detailed experimental results obtained with Columbia University COIL database. Finally, Section 6 summarizes the contributions of this paper.

2. BRIEF DESCRIPTION OF SIFT DESCRIPTORS

Recently, researchers have focused their exploration to local features in an image, which are invariant to image transformations and variations. Usually, there are two majors stages to found local features in an image. The first stage involves detecting features in an image in a repeatable way and the second one involves computing descriptors for each detected interest point.

An experimental evaluation of several different descriptors shows that the SIFT descriptors obtain the best results [3]. SIFT descriptors are reasonably invariant to changes in illumination, image noise, rotation, scaling, and small changes in viewpoint. Based on the robustness of SIFT descriptors described in [3], we choose SIFT approach to characterize ours images. SIFT as described in [10], consists of four steps. A brief description of SIFT method is presented here as mentioned below:

Step1: Scale-space peak selection

In the first step, possible interest points are identified by browsing the image over location and scale. This is implemented efficiently by constructing a Gaussian pyramid and searching for local peaks (termed keypoints) in a series of difference-of-Gaussian (DoG) images which can be computed from the difference of two nearby scaled images separated by a multiplicative factor *k*:

$$\begin{aligned} D(x,y,\sigma) &= \big(G(x,y,k\sigma) - G(x,y,\sigma) * I(x,y)\big) \\ &= L(x,y,k\sigma) - L(x,y,\sigma) \end{aligned} \qquad (1)$$

where $L(x,y,\sigma)$ is produced from the convolution of a variable-scale Gaussian, $G(x,y,\sigma)$, with an input image, $I(x,y)$:

$$L(x,y,\sigma) = G(x,y,\sigma) * I(x,y) \qquad (2)$$

The convolved images are grouped by octave. An octave corresponds to doubling the value of σ, and the value of k is selected so that a fixed number of blurred images are generated per octave. Keypoints are identified as local extrema (maxima or minima) of the $D(x,y,\sigma)$cross scales. In order to detect the local extrema, each pixel in $D(x,y,\sigma)$ image is compared to its 8 neighbors in the current image and 9 neighbors in the scale above and below. If a pixel is a local maximum or minimum, it is selected as a candidate keypoint.

Step 2: Keypoint localization

Once a keypoint candidate has been found by comparing a pixel to its neighbors, the next step is to filter so that only stable and more localized keypoints are retained. For each candidate keypoint, interpolation of nearby data is used to accurately determine its position. Then, keypoints with low contrast are removed and responses along edges are eliminated.

Step 3: Orientation assignment

In this step, an orientation is assigned to each keypoint localized in step 2. To determine the keypoint orientation, a gradient orientation histogram is computed from an orientation histogram of local gradients from the closest smoothed image $L(x,y,\sigma)$. For each image sample $L(x,y)$ at this scale, the gradient magnitude $m(x,y)$ and orientation $\theta(x,y)$ is precomputed using pixel differences:

$$m(x,y) = \sqrt{\left(\big(L(x+1,y) - L(x-1,y)\big)^2 + \big(L(x,y+1) - L(x,y-1)\big)^2\right)} \qquad (3)$$

$$\theta(x,y) = arctang\left(\frac{L(x,y+1) - L(x,y-1)}{L(x+1,y) - L(x-1,y)}\right) \qquad (4)$$

An orientation histogram is formed from the gradient orientations of sample points within
a region around the keypoint. The orientation histogram has 36 bins covering the 360 degree range of orientations. Each sample added to the histogram is weighted by its gradient magnitude and by a Gaussian-weighted circular window with a σ that is 1.5 times that of the scale of the keypoint. Peaks in the histogram correspond to dominant

orientations. A separate keypoint is created for the direction corresponding to the histogram maximum, and any other direction within 80% of the maximum value.
All the properties of the keypoint are measured relative to the keypoint orientation, this provides invariance to rotation.

Step 4: Keypoint descriptor
The local gradient data from the closest smoothed image $L(x, y, \sigma)$ is used to create the keypoint descriptor. This gradient information is first rotated to align it with the assigned orientation of the keypoint and then weighted by a Gaussian with sigma that is 1.5 times the scale of the keypoint. The weighted data is used to create a nominated number of histograms over a set window around the keypoint. Usual keypoint descriptors employ 16 orientation histograms aligned in a 4x4 grid. Each histogram has 8 orientation bins each created over a support window of 4x4 pixels. This leads to a SIFT feature vector with 128 elements with a support window of 16x16 scaled pixels.

3. THE PROPOSED APPROACH

Given an image, its keypoints are compared with keypoints of every image present in the database. Since the Euclidean distance has always been used to compare these keypoints so far, employing other approaches for the comparison is an interesting research direction that still needs to be explored. As mentioned in the introduction, we propose fuzzy similarity measures as alternatives for the Euclidean distance. This does not add any unwanted complexity because many fuzzy similarity measures are very easy to implement and compute, and fuzzy similarity measures offer the additional advantage of being studied extensively and having very solid theoretical foundations. Our hypothesis is that a similarity measure based on fuzzy measure should reflect a more objective similarity of content than classical distances.

Our approach is based on a fuzzy relation formalism described in section 3. In fact we measure the similarity by transforming each keypoints component of every image present in the database into fuzzy triangular number. These fuzzy measures allow to know the similarity degree between keypoints of images in the database. In the following, we explain how keypoints are compared using fuzzy similarities.

Formally speaking, our approach can be summarized as follows:
Given an *image A* containing an object and target image *B* possibly containing a similar object, the different steps used in our algorithm can be summarized as follows:

Step 1. Extract features from images *A* and *B* by applying SIFT descriptors:

SIFT transforms the image *A* and *B* into collection of local feature vectors represented by $\{A_{ik}, i = 1, 2, \dots, n; k = 1, 2, \dots, 128\}$ and $\{B_{jk},\ j = 1, 2, \dots, m; k = 1, 2, \dots, 128\}$. Each vector contains 128 components.

Step 2. Transform each component into fuzzy triangular number:

For each image and for each component; determine the fuzzy triangular number $I(k, A) = (a_1, a_2, a_3)$ as shown in Figure 1. So, the $I(k, A)$ represents the fuzzy number of the component *k* for image *A*, which *k* varies from 1 to 128. To determine the parameters a_1, a_2 and a_3, we based on the Chebyshev's theorem. Hence, for each feature from $\{A_{ik}, i = 1, 2, \dots, n; k = 1, 2, \dots, 128\}$ and $\{B_{jk},\ j = 1, 2, \dots, m;\ k = 1, 2, \dots, 128\}$
first we calculate the mean m_k and standard deviation σ_k for each image and for each component *k*.
We will use the mean and the standard deviation to determine the parameters a_1, a_2 and a_3 as follows:

$$a_1 = m_k - t\sigma_k,\ \ a_2 = m_k,\ \ a_3 = m_k + t\sigma_k \qquad (5)$$

for k= 1,... 128 and t is an integer number greater than 2.
With different experimental test, we have found that $t = 2$ is the best value.

In this work, we have adapted the *asymmetric* fuzzy triangular number described in [11] by using the fuzzy *symmetric* triangular number (see formula (5)). This is done in order to make easy the computation and the implementation of our method.

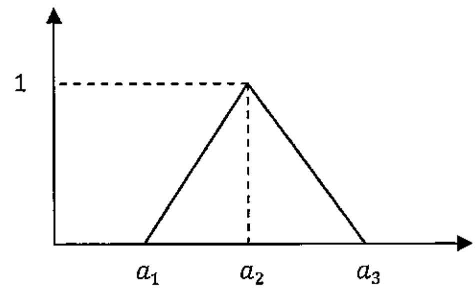

FIGURE 1: A FUZZY SYMETRIC TRIANGULAR NUMBER

Step 3. Measure the degree of similarity between images A and B:

Once the fuzzy triangular numbers are determined as in the previous steps, we calculate similarity degree between images A and B. The comprehensive similarity between images A and B is:

$$S(A,B) = \frac{1}{128}\sum_{k=1}^{128} S_k(A,B) \qquad (6)$$

The $S_k(A,B)$ is the partial similarity degree between images A and B on the component k. The partial fuzzy similarity represents the intersection surfaces between the fuzzy triangular numbers $I(k,A)$ and the $I(k,B)$. This intersection is calculated using the formalism described in section 4.
The approach described above has been implemented using MATLAB.

4. FUZZY SIMILARITY MEASURE

As explained in the end of the section 3, in order to measure similarity between images A and B, we need to measure partials similarities $S_k(A,B)$ where $k = 1,2,\dots,128$. A partial similarity $S_k(A,B)$ corresponds to the intersection surfaces between the fuzzy symmetric triangular numbers $I(k,A) = (a_1,a_2,a_3)$ and $I(k,B) = (b_1,b_2,b_3)$. An example of intersection is given in figure 2. From figure 2, we can notice that there are several possible cases between two symmetric triangles depending on the triangular numbers values. For each situation, there is an explicit formula which allows computing the intersection surface between two triangles. These formulas were inspired from the paper [11], there are defined in the following:

i. If $a_2 > b_2$:
In this situation, there are five possible situations between two symmetric triangles:

Case 1 : if $a_1 \geq b_3$, then $\big(I(k,A) \cap I(k,B)\big)=0$, since the two membership functions do not overlap;
Case 2 : if $(a_3 \geq b_3$ et $b_3 > a_1$ et $a_1 \geq b_1)$, then

$$\big(I(k,A) \cap I(k,B)\big) = \frac{(b_3-a_1)^2}{2.\big((a_2-b_2)+(b_3-a_1)\big)}$$

Case 3 : if $(a_3 \geq b_3$ et $b_1 \geq a_1)$, then

$$\big(I(k,A) \cap I(k,B)\big) = \frac{(b_1-a_1)^2}{2.\big((b_2-a_2)+(a_1-b_1)\big)} + \frac{(b_3-a_1)^2}{2.\big((a_2-b_2)+(b_3-a_1)\big)}$$

Case 4 : if $(b_3 \geq a_3$ et $a_1 \geq b_1)$, then

$$\big(I(k,A) \cap I(k,B)\big) = \frac{(b_3-a_1)^2}{2.\big((a_2-b_2)+(b_3-a_1)\big)} + \frac{(a_3-b_3)^2}{2.\big((b_2-a_2)+(a_3-b_3)\big)}$$

Case 5 : if $(b_3 \geq a_3 \text{ et } b_1 \geq a_1)$, then

$$\left(I(k,A) \cap I(k,B)\right) = \frac{(b_1-a_1)^2}{2.\left((a_1+b_1)-(a_2+b_2)\right)} + \frac{(b_3-a_1)^2}{2.\left((b_3-a_1)+(a_2-b_2)\right)} + \frac{(a_3-b_3)^2}{2.\left((a_3-b_3)+(b_2-a_2)\right)}$$

ii. If $b_2 > a_2$:
In this situation, we have the same cases and similar formulas like the situation i. discusses above. In formulas defined above, parameters $a_i, i = 1,2,3$ becomes $b_i, i = 1,2,3$ and vice versa.

iii. If $a_2 = b_2$:
In this situation, there are four possible situations between two symmetric triangles:
Case 1 : if $(a_1 \geq b_1 \text{ et } a_3 \leq b_3)$, then

$$\left(I(k,A) \cap I(k,B)\right) = \frac{(b_3-b_1)}{2}$$

Case 2 : if $(b_1 \geq a_1 \text{ et } b_3 \leq a_3)$, then

$$\left(I(k,A) \cap I(k,B)\right) = \frac{(a_3-a_1)}{2}$$

Case 3 : if $(a_1 \geq b_1 \text{ et } a_3 \geq b_3)$, then

$$\left(I(k,A) \cap I(k,B)\right) = \frac{(b_3-a_1)}{2}$$

Case 4 : if $(a_1 \leq b_1 \text{ et } a_3 \leq b_3)$, then

$$\left(I(k,A) \cap I(k,B)\right) = \frac{(a_3-b_1)}{2}$$

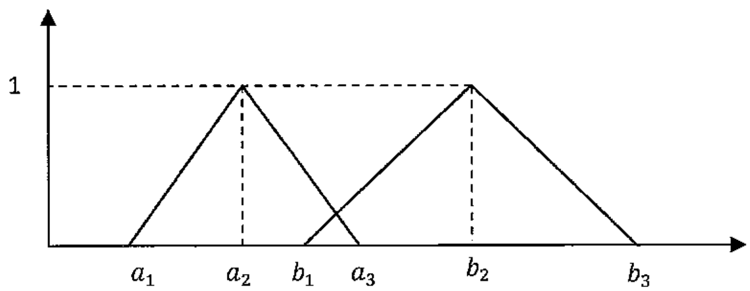

FIGURE 2: SAMPLE OF INTERSECTION BETWEEN TWO SYMMETRIC FUZZY TRIANGULAR NUMBERS

5. EXPERIMENT RESULTS

We have tested the proposed approach on Columbia University COIL database. COIL database consists of 100 objects recorded under 72 different viewing angles. In figure 3, sample of objects from this database are shown. In order to identify keypoints, SIFT method was applied to each image of the database. In figure 4, shows an image with the extracted keypoints. These keypoints are used in order to measure similarity between images with the proposed method. The experiment results on several images show that in combining SIFT descriptors and fuzzy similarity, our approach was able to recognize a single object. As a preliminary evaluation, figure 5 presents an example of our recognition results considering the top-left image the object to be recognized. The other images are the ones found in descending order from the left to right then top to bottom. In fugure 4, we notice that the method found several correct images. However, if the viewing angle is extreme, the proposed approach fails to recognize an object in an image. We are presently investigating in more details the combination of learning algorithms and color information to our approach. This is in order to improve the performances and also to apply enhanced approach to other Content Based Image Retrieval in real world problems from medical image diagnosis and face recognition.

FIGURE 3: SAMPLE VIEWS FOR THE OBJECTS (COLUMBIA UNIVERSITY COIL DATABASE)

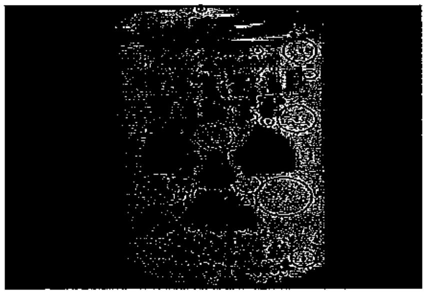

FIGURE 4: AN IMAGE WITH THE EXTRACTED KEYPOINTS

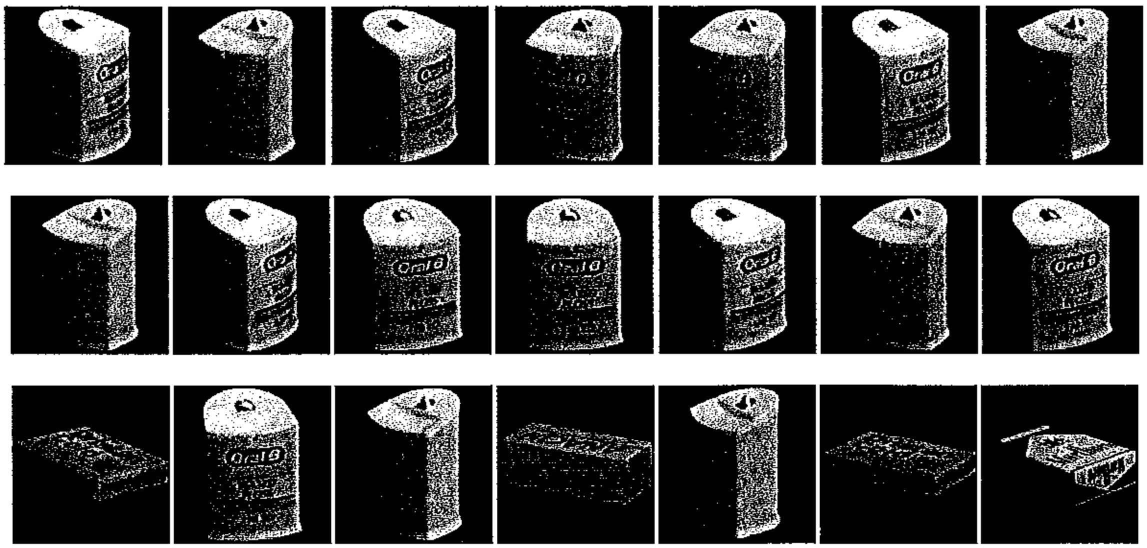

FIGURE 5: EXAMPLE OF RECOGNITION

6. CONCLUSION

We have presented a method for object recognition based on SIFT descriptors and fuzzy similarity measure. The proposed method is based on two main components. First, keypoints are extracted from images using SIFT descriptors. Then, similarity measure is computed between SIFT descriptors applying fuzzy similarity measure. The method performs well in the case of single objet in an image. From the results obtained we have shown that fuzzy similarity with SIFT have significant potential in object recognition. In our knowledge, it is the first time the SIFT descriptors was used with fuzzy to recognize objects. Future work with, we would like to work on learning algorithms in order to improve the performances. This step is essential because viewing angle reduces significantly the performance of the proposed approach.

7. ACKNOWLEDGEMENTS

We would like to thank David Lowe for the SIFT feature code. This work is supported by NSERC, Canada.

REFERENCES

[1] Ke, Y., Sukthankar, R. and Huston, L., "An efficient parts-based near-duplicate and sub-image retrieval system," In Proc. Int. Conf. on Multimedia, 869–876 (2004).

[2] Tuytelaars, T., and Gool, L.V., "Content-based image retrieval based on local affinely invariant regions," In Proc. Int. Conf. on Visual Information and Information Systems, 493–500 (1999).

[3] Mikolajczyk, K., and Schmid, C., "Indexing based on scale invariant interest points," In Proceedings of International Conference on Computer Vision, 525–531, (2001).

[4] Fergus, R., Perona, P., and Zisserman, A., "Object class recognition by unsupervised scale-invariant learning," In Proceedings of Computer Vision and Pattern Recognition, (2003).

[5] Lowe, D. G., "Object recognition from local scale-invariant features," In Proceedings of International Conference on Computer Vision, 1150–1157 (1999).

[6] Brown, M. and Lowe,D. G., "Invariant features from interest point groups," In British Machine Vision Conference, BMVC 2002, Cardiff, Wales, 656-665 (2002).

[7] Lazebnik, Schmid, S. C. and Ponce, J., "Affine-invariant local descriptors and neighborhood statistics for texture recognition," In International Conference on Computer Vision, (2003).

[8] Yu, J., Tian, Q., Amores, J. and Sebe, N., "Toward Robust Distance Metric Analysis for Similarity Estimation," Proc. of IEEE CVPR, vol. 1, 316-322 (2006).

[9] Yu, J., Amores, J., Sebe N. and Tian,Q., "A New Study on Distance Metrics as Similarity Measurement," Proc. of IEEE ICME, (2006).

[10] Lowe, D. G., "Distinctive image features from scale-invariant keypoints," International Journal of Computer Vision, 60, 2, 91-110 (2004).

[11] Cheng-Jian Lin and Wen-Hao Ho "An asymmetry-similarity-measure-based neural fuzzy inference system" Fuzzy Sets and Systems, Elsevier, 152, 535–551 (2005).