

# MGS-SIFT: A New Illumination Invariant Feature Based on SIFT Descriptor

Reza Javanmard Alitappeh and Fariborz Mahmoudi

**Abstract**—The SIFT descriptor is one of the most widely used descriptors which has considerable stability against changes such as rotation, scale, affine of the image, and illumination. However, because of the greater emphasis on its insensitivity to geometric changes, this descriptor is weak in various illuminations. Therefore, in this article an attempt has been made to boost the SIFT descriptor against changes in illuminations through the use of techniques of creating pictures in various illumination conditions and by extracting the desired features of these conditions. For this purpose, we have used the Power-Law Transform, and the results of the implementation testing have been successful. The efficiency of the proposed algorithm and of the base algorithm of SIFT with regard to the data set ALOI have been investigated, and it has been found that by adding this method to the base SIFT descriptor, the rate of recognition improves by five percent. Moreover, there will be a better response to changes in illumination.

**Index Terms**—Illumination variance, object recognition, power-law transform, SIFT descriptor.

## I. INTRODUCTION

Extracting the key points from the image of an object (i.e., points which can function as good representatives for describing the object) which are stable in various views and make the realization of good recognition possible, is one of the main challenges in the area of machine vision, camera calibration, 3D reconstruction, image registration, robot navigation, and object recognition are only a few of the applications of these features. For example, in object recognition these key points can be used in three stages:

- 1) Finding the key points: this can be achieved through a general strategy by searching in the length of the image and finding unique points which have specific features. These points can be found by searching for the corners, bubbles, and T-junctions.
- 2) Describing the key points: these points should be described in a way that we will have the same display of the key points in the presence of noise in the environment and when changes occur in geometry and illumination, and that these points will be distinctive and insensitive.
- 3) The last stage concerns the matching of these points among different images. Normally, methods of

Manuscript received June 27, 2012; revised August 17, 2012. This work was supported in part by the Mechatronic Researcher Laboratory (MRL) Qazvin, Iran. MGS-SIFT: A New Illumination Invariant Feature Based on SIFT Descriptor.

Reza Javanmard Alitappeh and Fariborz Mahmoudi are with Islamic Azad University Qazvin, Iran. (e-mail: R\_Javanmard@b-just.ac.ir, Fzmahmoudi@gmail.com).

calculating distances, such as Euclidean and Mahalanobis methods are used for eigenvectors obtained in the last stage. In a comparison carried out among different methods of describing features [1], it was found that the insensitivity to scale transform in SIFT [2], [3] offers the most distinctive description of the object. The SIFT descriptor emphasizes on key insensitive points extracted through Gaussian Differences (DoG). In the description stage, the magnitude and orientation of the images, gradients based on histogram, and gradient orientations around the key points are also obtained.

This descriptor yields good results in changes such as rotation, scale, and affine of the images, but it is weak against changes in illumination. Therefore, in this article a method with the name MGS1 is presented to overcome this problem through the use of techniques of constructing new samples of an image in different illumination conditions. The innovation introduced in this article is the collection of more points from the image in different illumination conditions (Fig. 1.) in order to help the image become independent of lighting. Adjustment of parameters has been carried out through the use of genetic algorithms. One of the advantages of the proposed method is that most of the processing is performed offline. Therefore, there will be little overhead when the testing is done; and the insensitive recognition of illumination will be performed online. The practical tests carried out in this article show that the proposed method had a considerable effect in increasing the number of matched points and it also increased the accuracy of classification.



Fig. 1. Different illumination of an image [14]

The rest of the article is organized as follows: in part two previous studies related to recognition of objects, in particular with the aim of making the descriptors insensitive to illumination, are described. In part three, the base SIFT descriptor, along with details of our method, is explained. The proposed method is discussed in part four. And finally, the implementation results are presented.

## II. RELATED WORK

As was stated before, strong local descriptors, which are

<sup>1</sup> - Multi Gray Scale

obtained through extracting key insensitive points, have had many applications in image recovery, camera adjustment, object recognition, etc. during the last two decades.

Among the applications presented up to now, it can be said that the SIFT descriptor has had the best results in the rate of recognition. Therefore, in the rest of the article different versions of this descriptor are reviewed. In each version, we have tried to boost one of the features.

We have also added some new features to some of the versions. For example, we have used the PCA-SIFT [4] method to reduce the dimensions of the eigenvector of the base SIFT from 128 to 36.

SIFT, in its general sense, adds the features included in figure to the previous eigenvector [5] in order to raise the distinguishing power at times when there has happened a similar contextual construction in the images. Michel and Yu introduced the descriptor ASIFT in 2009 [6]. This descriptor, besides having all the features of SIFT, has a very great ability for pictures which have experienced affine transform.

The CSIFT descriptor [7] is the first method which has taken the feature of color into consideration. This descriptor, acting on the idea that the feature of color contains useful information, has added the model insensitive to color [8] to the eigenvector of SIFT.

Of course, this has resulted in a greater calculation load and in more complexity for transforms such as scale, rotation, etc. For this reason, this descriptor lacks the desired efficiency in some cases. Since color plays an important role in the description and recognition of objects and provides scholars in the area of machine vision with information such as image histogram [9], colored insensitive moments, and co-occurrence matrix [10], etc.

Other studies were also suggested to improve performance. For example, in a different method in CSFIT [11], the co-occurrence matrix has been used in introducing the features of color in the descriptor SIFT. The initial co-occurrence matrix could be used on images having a grey surface, but other versions of it can be used on colored images as well. This method is known as SIFT-CCH because these two features are combined in one eigenvector.

It can be seen that the above mentioned methods have been introduced for different purposes. For example, the last two methods are used when we have two completely identical images in two different colors. However, in this article we have tried to boost the base SIFT feature at times when there are different MSGs of one image. In other words, one of the shortcomings which were not paid much attention to in the above mentioned studies is the insensitivity of the descriptor to different MSGs. With the purpose of achieving this goal, we have used the Power-Law Transform [12] in the construction of new samples of the main image with different MSGs. Then we have applied the SIFT feature on these samples and found out that the model constructed by using this transform had high resistance to changes in MSG. Of course, to reduce the time needed for calculations, we constructed a single image from the images obtained in the Power-Law Transform and all comparisons were carried out on this single final image.

### III. BASE SIFT DESCRIPTOR

Since the innovation we present is the improvement of the SIFT descriptor under conditions of different MSGs, in this section we discuss this descriptor. Today, the SIFT descriptor is considered as one of the best and most powerful tools for extracting key points insensitive to different conditions such as rotation, scale, change in viewing orientation, noise, MSG, and the affine transform. As was stated before, the SIFT descriptor has many advantages over other descriptors, and for this very reason it has received a lot of attention in the area of object recognition. In this application, through matching the key points extracted from the original image with their equivalent points in the final image, and by considering a given number of matched points, recognition is performed. In other words, it can be said that this descriptor does not learn the general features of an object in order to classify it. In one of the applications, nowadays SIFT is implemented in some of the FPGA boards. In general, this descriptor is used in three stages which are as follows:

#### A. Finding the Key Point

The first stage in all methods in which work is carried out on special (key) points of the picture is to find these points.

##### 1) Finding extreme points in scale space

In this method, Gaussian Differences (DoG) are used to find the key points of the image. The process of finding these points starts with constructing a pyramid of images, and with convolution of the image  $I(x,y)$  with a Gaussian filter  $G(x,y,\sigma)$ . Therefore, the scale-space is represented as follows:

$$L(x, y, \sigma) = I(x, y) * G(x, y, \sigma) \quad (1)$$

"\*" stands for the convolution operator in  $(x, y)$ , and

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \quad (2)$$

The degree of blurring is controlled by the standard deviation parameter  $\sigma$  in the Gaussian function.

The scale-space DoG can also be obtained by subtracting adjacent scale levels:

$$D(x, y, \sigma) = [G(x, y, k\sigma) - G(x, y, \sigma)] * I(x, y) \quad (3)$$

Using the Eq. 1 we will get:

$$D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma) \quad (4)$$

In the next stage, the maximum and the minimum points in each octave are found. This is achieved by comparing each pixel with its 26 neighbors in region  $3*3$  of all adjacent DOG levels present in the same octave (Fig. 3.). If the point in question is bigger or smaller than all its neighbors, it is chosen as the desired point [3].

##### 2) Locating Key Points

In this stage, we omit some of the extracted key points in two phases so that we can have key points which are less sensitive to noise, or have points which are not located on the edges. To do this, in phase one we use the Taylor expansion to omit extreme points which are unstable and have a low Power-Law contrast [13]. Since DoG has high sensitivity to edges, it is possible that some extracted points will be along weak edges and hence will not be stable in the

presence of noise. Therefore, in phase two we use the Hessian matrix to omit points which have the above feature.

### 3) . Orientation Assignment

In this stage, preparations are made for constructing the eigenvector. To each key point, an orientation is assigned based on the local features of the image. For each sample image  $L(x, y)$  in this scale, the gradient range  $m(x, y)$  and the gradient orientation  $\theta(x, y)$  are calculated using differences in pixels and based on the following formulae:

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2} \quad (5)$$

$$\theta(x, y) = \tan^{-1}((L(x, y+1) - L(x, y-1)) / (L(x+1, y) - L(x-1, y))) \quad (6)$$

In Fig. 2. the stages of constructing the DoG space are shown.

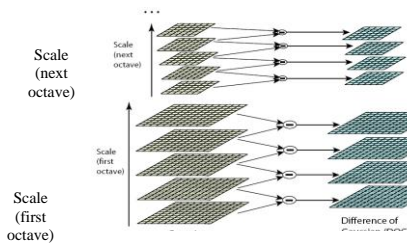


Fig. 2. For each octave of scale space, the initial image is repeatedly convolved with Gaussians to produce the set of scale space images shown on the left. Adjacent Gaussian images are subtracted to produce the difference-of-Gaussian images on the right. After each octave, the Gaussian image is Down-sampled by a factor of 2, and the process repeated.

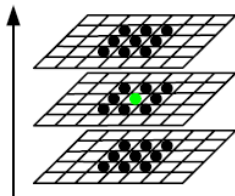


Fig. 3. Maxima and minima of the difference-of-Gaussian images are detected by comparing a pixel (marked with X) to its 26 neighbors in 3x3 regions at the current and adjacent scales (marked with circles).

The orientation histogram is built by using the gradient range for a key point together with an area around the point (Fig. 4.)

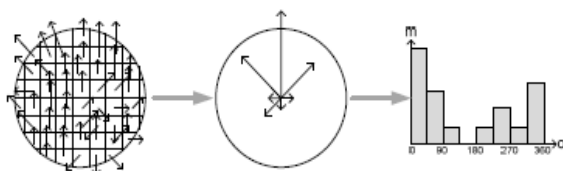


Fig. 4 (a) Image gradient (b) Key point descriptor (c) Orientation Histogram

### B. Display the Key Point Description

In this stage the eigenvector will be developed. First the gradient range and the orientation around the key point are sampled. In his experiment, David Lowe used the  $4 \times 4$  array with eight orientations in each histogram instead of

employing a  $2 \times 2$  array for orientation histograms. Therefore, the eigenvector length will be  $4 \times 4 \times 8 = 128$  members for each key point.

### C. Feature Vector Matching

The matching phase of the recognition stage is carried out by comparing each key point extracted from the test image with key points related to the trained image. The best candidate points for matching are found through recognizing the nearest neighbor in the set of key points of the train image. The nearest neighbor has the least distance from its matched point.

## IV. PROPOSED ALGORITHM

The following flowchart shows the procedure of carrying out the processes of the proposed algorithm.

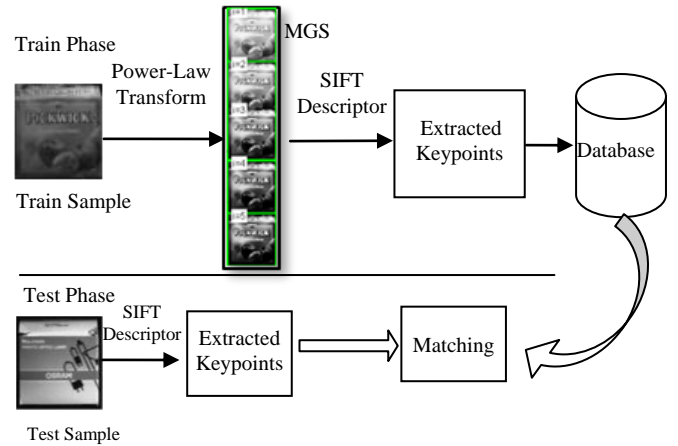


Fig. 5. Flowchart of proposed algorithm

After applying the Power-Law Transform on the train sample, the key points are extracted from the MGS (Multi Gray Scale) space by the SIFT descriptor and are stored in the database of the train set. Care must be taken that the above process is carried out offline. In the testing phase, after extracting the key points from the testing sample, the operations matching the key points of this sample with all the samples present in the database are performed. The nearest sample, which has the most matched points, is introduced as the chosen class.

### A. Power-Law Transform

The Power-Law Transform is defined as follows:

$$s = cr^\gamma \quad (7)$$

where  $c$  and  $\gamma$  are positive constants. Sometimes Eq. 7 is written as  $s = c(r + \epsilon)^\gamma$  when relocation is considered (with the condition the entry be zero). However, in Eq. 7 we have omitted the relocation value, because we have assumed that the conditions are normal and that calibration has been carried out. Changes in the two variables  $r$  and  $c$ , with respect to each other, are shown in Fig. 6. for various values of  $\gamma$ .

Power-Law curves having a small  $\gamma$ , map an area of dark input value in an extensive area of output values, the reverse is also true for levels with higher input values. Different and possible states of the transform curves can be easily obtained by changing the value of omega.

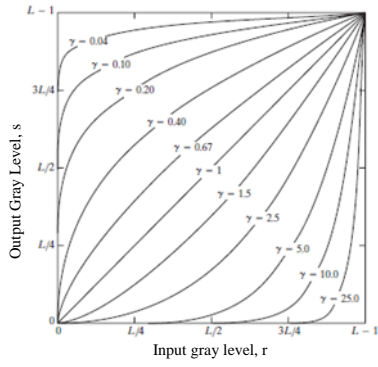


Fig. 6. Power-Law transform for different value of  $\gamma$

Based on Fig. 6. it can be seen that the curve relating to values from  $\gamma > 1$  are the converse of the curves produced by values from  $\gamma < 1$ . The last point to find out from Eq. 7 is the simultaneous transform with  $c = \gamma = 1$ . Various image capturing, printing, and displaying equipment operate based on this transform. The defining parameter in this equation is gamma, and the process which makes the use of this transform suitable for a particular application is called gamma correction. Since this transform has a variety of applications, the gamma correction has a very important role in improving the efficiency of the output. It is for this very reason that the optimum values of gamma have been calculated using genetic algorithms.

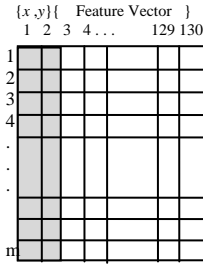


Fig. 7. Key points with their feature vectors for a sample like  $s_i$

**B. Using SIFT to Key Point Extraction**

Assuming that there is a set of objects  $S$  in the form of  $S = \{s_1, s_2, s_3, \dots, s_n\}$ , for each  $s_i$  form the set  $S$ , there is a set of key points with their eigenvectors that have been extracted by SIFT (Fig. 7). In other words, each row in this table represents one of the key points in our train image.  $m$  shows the number of key points related to an image.  $(x, y)$  are the coordinates of the key point.

In the testing phase, which is usually in real-time, first the key points together with eigenvectors are extracted from the test image. Therefore, we will have two sets, one representing the train sample and the other related to the test set. We compare the key points of the test set with each and every row in the  $S$  set, and if the number of matched points for each  $s_i$  is more than the others, that member is determined as the matched object (Fig. 8).

**C. Combination of SIFT and Power-Law Transform**

A closer look at the proposal algorithm will show that it consists of the following stages:

- 1) Constructing samples insensitive to illumination (MGS) using the train samples ( $S$ ).
- 2) Extracting key points and extracting feature vectors

- 3) Transforming the new samples from the MGS space of each class ( $s_i$ ) to a single sample ( $s_{ij}$ )
- 4) Continuing the algorithm based on the base SIFT descriptor.

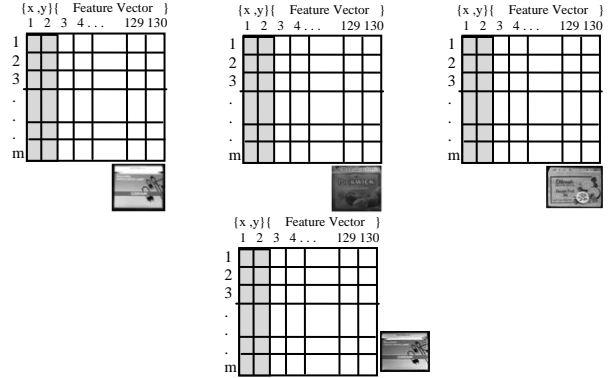


Fig. 8. Feature vector representation for train set (top), test set (bottom)

In the first stage, after applying the above transform on the set of train data set  $S$ , the set  $S = \{s_{1,1}, s_{1,2}, \dots, s_{1,k}, s_{2,1}, \dots, s_{2,k}, \dots, s_{n,1}, \dots, s_{n,k}\}$  or MGS, is produced. The  $s_{ij}$  ( $i=1..n, j=1..k$ ) represents the  $i$ th sample having the  $j$ th illumination level. (In our tests, we have chosen the value of 5 for  $k$ ).

In the second stage, the key points and their eigenvectors are extracted from MGS. To reduce the computation load, the key points present in the MGS ( $s_{11}$  to  $s_{1k}$ ) are transformed to a single final image ( $s_{ij}$ ), in a way that this final image contains the features of all the key points in various illuminations. Therefore, the final  $S$  set will be in the form of  $s = \{sf_1, sf_2, \dots, sf_n\}$

There are two strategies for combining key points in the MGS into the one image (Fig. 9.): either only common points (rectangle points) are chosen or common key points and key points which are not common are chosen (circle points). Were observed in experiments the second approach have better performance with more different.

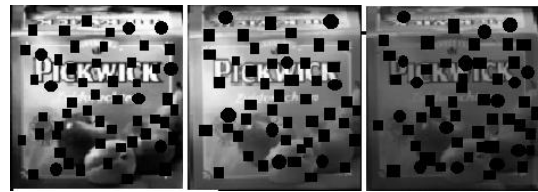


Fig. 9. MGS space with common points (Rectangle points) and not common points (Circle points)

**D. Gamma Correction with Genetic Algorithm**

TABLE I: PARAMETER VALUES USED IN GENETIC ALGORITHMS

TrainSet	$\alpha$	K	High Rang	Low Rang	$P_m$
20-70%	80%	5	10	0	0.5

$P_m$  mutation rate, Low/High range Gens confine parameters (admissible confine of  $\gamma$ ), K number of scales of MGS space,  $\alpha$  parameters division as fitness impact factor (Eq. 8) and the size of train set appointed with TrainSet.

$$TotalFit = \alpha \times ClassificationFit + (1 - \alpha) DisperalFit \quad (8)$$

$$\text{DispersalFit} = \sum_{i=2}^k \text{dist}(x_i + x_{i-1}) \quad (9)$$

In fitness computation,  $\alpha$  in recognition rate and  $1-\alpha$  in rate of dispersal are multiplied. Scattering rates in order to maintain the distance between various  $\gamma$ .

## V. EXPERIMENTAL RESULT

The data set used in the tests ALOI [14] included 27 samples with different illumination for 1000 different objects. (Fig. 1 and 9 are show that)

In the first test, adaption points compared with two methods. Points that adapted via MSG-SIFT algorithms are more than of this point that obtained with SIFT base. (Fig. 10) Therefore, when the illumination conditions in different data sets we tested, the proposed algorithm will be more reliability.

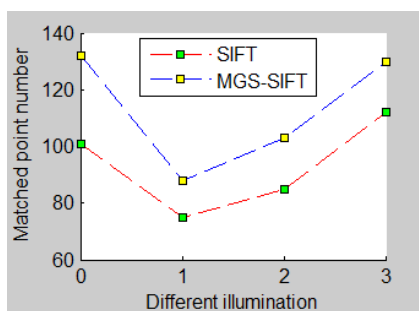


Fig. 10. Comparing SIFT vs. MSG-SIFT. matched key point number (vertical axes), 4 different illumination (horizontal axes)

In the second experiment the number of different training and test samples was considered. As the result of experiments in Table II shows, in the absence if the number of samples is still small given the proposed algorithm has higher accuracy will be. This is because the distinctive features in the construction of new samples with different illumination conditions.

TABLE II: EXPERIMENTAL RESULTS. COMPARING BASE SIFT DESCRIPTOR vs. MSG-SIFT

Train and test set percent/Approach	SFIT	MGS-SIFT
20% Train,80% Test	85.77	89.64
70% Train,30% Test	89	95

So you can conclude that the proposed algorithm on a data set containing examples of the lighting is different is a better result. Therefore, in such matters, the algorithm has provided will be better performance than based SIFT.

## VI. CONCLUSION

Two Parameters that should be satisfied by the descriptor is stability and distinction. (Repeatability against changes and having the minimum information to describe objects)

As observed, the proposed method in this article, one of the most popular descriptors in the field of machine vision stability and distinction in two aspects examined.

With regards to the conversion descriptor SFIT rotation, scaling and image stretching and strengthening is invariant against changes in illumination certainly more issues will be used. But in later work on this issue can be reviewed and the color images in color space problem can be solved.

## ACKNOWLEDGMENT

This project is supported with Mechatronic Research Centre of Qazvin Azad University.

## REFERENCES

- [1] K. Mikolajczyk and C. Schmid, "Scale and Affine Invariant Interest Point Detectors," *International Journal of Computer Vision*, vol. 1, no. 60, pp. 63-86, 2009.
- [2] D. Lowe, "Object Recognition from Local Scale-Invariant Features," in *Proc. of 7th International Conference on ComputerVision*, vol. 19, pp. 15-17.
- [3] D. Lowe, "Distinctive Image Features from Scale percent Invariant Keypoints," *International Journal of Computer Vision*, vol. 2, no. 60, pp. 91-110, 2004.
- [4] E. Mortensen, H. Deng, and L. Shapiro, "A SIFT descriptor with global context," in *Proc. of Conf. on Computer Vision and Pattern Recognition*, 2005.
- [5] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *WEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 4, pp. 509-522, 2002.
- [6] J. M. Morel, and G. Yu, "ASIFT: A New Framework for Fully Affine Invariant Image Comparison," *SIAM J. IMAGING SCIENCES*, *Society for Industrial and Applied Mathematics*, vol. 2, no. 2, pp. 438-469, 2009.
- [7] A. E. A.Hakim and A. A. Farag, "CSIFT: A SIFT Descriptor with Color Invariant Characteristics," in *Proc. of Computer Vision and Pattern Recognition Conf.* 2006, pp. 1978- 1983.
- [8] J. M. Geusebroek, R. V. D. Boomgaard, A. W. M. Smeulders, and H. Geerts, "Color Invariant," *IEEE Transaction on Pattern Analysis, and Machine Intelligence*, vol. 23, no. 12, pp. 1338-1350, 2001.
- [9] M. Swain and D. Ballard, "Color indexing," *International Journal of Computer Vision*, vol. 7, no. 1, pp.11-32, 1991.
- [10] S. O Shim and T. S Choi, "Image Indexing by Modified Color Co-occurrence Matrix," in *Proc. of IEEE International Conference on Image Processing*, vol. 3, pp. 493-496, 2003.
- [11] C. Ancuti and P. Bekaert, "SIFT-CCH: Increasing the SIFT distinctness by Color Co-occurrence Histograms," presented at 5th International Symposium on image and Signal Processing and Analysis, 2007.
- [12] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 2nd Edition, 2002 by Prentice-Hall, Inc. Upper Saddle River, New Jersey 07458.
- [13] M. Brown, and D.G. Lowe. "Invariant features from interest point groups." In *British Machine Vision Conf., Cardiff, Wales*, 2002, pp. 656-665.
- [14] J. M. Geusebroek, G. J. Burghouts, and A. W. M. Smeulders, "The Amsterdam library of object images," *International Journal of Computer Vision*, vol. 61, no. 1, pp. 103-112, 2005.



**Reza Javanmard Alitappeh** was born in Behshahr, Mazandaran, Iran in 1985. He received his Bsc. Degree in Computer Engineering- Software from Shomal University of Amol in Jul-2005, and in Jan-2011 his MSc in Artificial Intelligence from Azad University of Qazvin, Iran. Reza has joined to Mechatronic Researcher Laboratory (MRL) during his Master. And he has worked as @Home robot software/team leader till 2011. His research interest is Machine Vision, Machine Learning and Robotics and in this way he published many papers in journals and conferences.



**Fariborz Mahmoudi** was born in Tehran, Iran in 1966. He received post diploma in electronic in 1986 from Amirkabir University of Technology Iran, B.E. degree in computer software engineering in 1991 from Shahid Beheshti University Iran, M. Sc. degree in computer engineering in 1994 from Amirkabir University of Technology Iran (Poly-Technique Tehran) and Ph.D. degree in computer engineering from the Science and research campus of Azad University Iran. Since 1996, he has been a lecturer of computer engineering department at the Qazvin Azad University, Qazvin, Iran. As a senior researcher too, he was in the Iran Telecomm Research Center (Tehran, Iran) between 2002 and 2006, and in the Mechatronics Research Lab of Qazvin Azad University from 2006 till now. His research interests include machine vision, machine learning and information retrieval in the image and text databases. Dr. Mahmoudi has published more than 100 papers in the international scientific journals and conference proceedings.