Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 79 (2016) 543 – 552

Procedia Computer Science

7th International Conference on Communication, Computing and Virtualization 2016

FACE RECOGNITION USING FILTERED EOH-SIFT

Vinay A, Ganesh Kathiresan, Durga Akhil Mundroy, H. Nihar Nandan, Chetna Sureka, K N Balasubramanya Murthy, S Natarajan

Department of Computer Science and Engineering, PES University, 100 Feet Ring Road, BSK III Stage, Bengaluru, Karnataka 560085

Abstract

This paper presents a method for the implementation of facial recognition using filtering techniques that will increase the accuracy of the process as well as distinguish faces more decisively. The process has proved to be invariant to image scale, rotation and illumination. This paper was motivated by the EOH-SIFT approach. The process described in this paper is aimed at improving upon the effectiveness of EOH-SIFT by feeding it a filtered image. After much exhaustive research into various filters, this paper shows that, with two filters which when used in a specific order, significantly boost the potency of the EOH-SIFT approach to identify faces. This approach has given very promising results when tested on the ORL database. Although it is a standard dataset, it does not have much variation as seen in the real world. Hence, our approach has been tested on other datasets and it has produced extremely encouraging results. The recognition of faces proceeds by first obtaining the region of interest and applying the filters on that area followed by the identification of important features in the faces and then matching them using an efficient nearest-neighbor algorithm. This leads to a robust and definitive face recognition system.

Peer-review under responsibility of the Organizing Committee of ICCCV 2016

Keywords: Preprocessing, Filtering, Image feature point descriptors, Illumination Invariant, SIFT, EOH

1. Introduction

Face recognition is a person identification system which uses various digital image processing and pattern recognition techniques. Facial recognition has a wide range of applications and is commonly used for security purposes, categorizing photo galleries in various social networking sites and even personal collections. For example, faces of suspects from CCTV camera footage can be cleaned and analyzed against a dataset to get a positive identification like the system used in the Tocumen International Airport in Panama. Systems like this can only prove

Peer-review under responsibility of the Organizing Committee of ICCCV 2016
doi:10.1016/j.procs.2016.03.069

useful where manual checking can cause a slowdown or hinder the activities in that area. In countries like Mexico, facial recognition has been used to prevent voter fraud which has shown good results[4]. Countries like Australia and New Zealand also have implemented facial recognition in their border custom services system called "SmartGate" which compares captured photograph to the photograph in the e-passport microchip so as to verify the owner of the passport. Applications like Google's Picasa can automatically tag various people by using face recognition in combination with other machine learning techniques. These use user tagged photos to improve accuracy of the recognized face. Also, face detection and expression recognition can be used in various situations. In the United States, Law enforcement agencies and forensic investigators use facial recognition to map faces in pictures to an ID. Also, in the US, facial recognition is used to process the large number of visa applications and they operate on an extremely large scale using a database which contains in excess of 75 million photographs. There are many more use cases for facial recognition like identifying account holders who use the ATM where the photo of the person using the ATM would be compared to a photo already in the bank's database. Modern phones now incorporate facial recognition in their lock screens to securely identify the user of the device. There are a few devices that help people diagnosed with prosopagnosia to recognize people they know. Facial recognition is not only used to identify a person, but it is also used to get a lot of personal information about the test subject like blog posts, social networking accounts and other patterns associated with their photograph. This can be used extensively in the crime investigation sector. Digital cameras also incorporate face detection and emotion recognition to help focus the image on specific areas of a scene. The main advantage of using facial recognition over other biometric techniques is that there is no need for the co-operation of the test subject. Also, this system can be implemented on a large crowd of people at once. In such situations, conventional biometric techniques cannot be used to achieve a practical advantage. Since facial recognition has proved to be a very important topic of discussion, a lot of research has been done and one of the most widely used algorithms to make facial recognition systems is SIFT [1]. The approach described in this paper, improved upon EOH-SIFT [5] to increase its accuracy by using a combination of filters.

2. Variants in SIFT

2.1. PCA-SIFT

Like SIFT, this descriptor encodes the salient aspects of the image gradient in the feature point's neighborhood; however, instead of using SIFT's smoothed weighted histograms, it applies Principal Components Analysis (PCA) [2] to the normalized gradient patch. PCA-based local descriptors are more distinctive, more robust to image deformations, and more compact than the standard SIFT representation.

2.2. CSIFT

General SIFT will work with gray scale images. Typically, color provides vital information for image description and matching. Abdel-Hakim et. al. [3] addressed this problem by developing a new colored invariant SIFT. CSIFT has proved to be more robust than the conventional SIFT with respect to color and photometrical variations. The experimentation conducted on various datasets supports the potential of the proposed approach.

2.3. GSIFT

GSIFT[6] binds Global information into SIFT. The main objective is the addition of global texture information in order to make the descriptor include a wide range of curve shape information. For each keypoint detected, it establishes a vector with two parts. One is the SIFT descriptor of a local feature. Another is a global texture vector which can be used to distinguish similar local features.

2.4. ASIFT

SIFT cannot be used on images with affine changes. For better adaptability, ASIFT [9] initiates the rotation of camera's optical axis. Rotation transformation is first applied on an image and then tilt transformation, which is applied on its output which are got by changing the longitudinal and the latitude angle within a certain range. Keypoints are detected and which help establish the description from the affine image. In contrast to SIFT, ASIFT detects more feature points with a fewer mismatching points.

2.5. KPB-SIFT

Invariant feature descriptors are generally of high dimensionality like 128 dimensions in the case of SIFT. This limits the scale and speed of its performance. Thus another compact feature descriptor called Kernel Projection Based SIFT (KPB-SIFT) [8] was formulated. Kernel projection techniques are applied to orientation gradient patches unlike the case of SIFT's smoothed weighted histograms. This makes KPB-SIFT more compact, and shows advantages in terms of distinctiveness, tolerance to geometric distortions and invariance to scale to better effect.

3. Proposed Approach

This paper proposes to use the cascade object detector to extract the region of interest and then, use a combination of image filters on the extracted region to pre-process the image so as to make the EOH-SIFT algorithms more accurate. After much trial and error and testing of all image filtering techniques a conclusion can be reached that the combination of medfilt and rangefilt in that specific order gave more accurate results. Hence, this paper explores the reasons for this effect on the accuracy of the EOH-SIFT results.

3.1. Image Acquisitions

To test our thesis, the following two datasets which were generously made available by various organizations have been used.

3.1.1 FEI Dataset.

This dataset of faces is a Brazilian face database that contains a set of pictures taken between 2005 and 2006. This dataset was made by the students and staff at the Artificial Intelligence Laboratory of FEI in São Bernardo do Campo, São Paulo, Brazil. 200 persons of a large age range have posed for this dataset and it contains 14 images for each of the 200 individuals, a total of 2800 images. Scale might vary about 10% and the original size of each image is 640x480 pixels. All images are col-our pictures and are taken against a white homogenous background in an upright frontal position with profile rotation of up to about 180 degrees. The age ranges be-tween 19 and 40 years old with distinct appearance, hairstyle, and adorns. This dataset has been chosen as not only are the facial features very clear and highly distinct, but also the pictures have appropriate resolution. This has made them easier to process and helped us to get a more error free result. The dataset also contains sufficient images of each individual which has resulted in us testing them in many ways. Due to this dataset containing a wide range of ages, it gave us a fair chance to test the validity and accuracy of the algorithm.

3.1.2 CALTECH Dataset.

This dataset was collected and organized by Markus Weber at California Institute of Technology. It is frontal face dataset that has around 450 images of 27 people whose photographs were taken under different backgrounds, expressions and lighting conditions. This dataset was chosen as it contains photos with a wide range of expressions which can tend to impact the algorithm used in a big way. Also, the wide range of backgrounds and high resolution of the images enabled us to rigorously test our approach.

3.2. Region of Interest

Region of Interest (often referred to as just ROI), is a subset of selected samples from a dataset for a particular purpose. This concept of ROI is used in many areas for example in medical imaging, for studying a tumor, the boundaries maybe defined on an image or in a volume. In face detection, Region of Interest (ROI) is the portion of an image that you want to perform filter or some other operations on. ROI can be defined by creating a binary mask, which is an image of the same size as the one you want to perform operations on. In the binary mask, the pixels are set to 1 that defines the image and all other pixels set to 0.Multiple ROI can be initialized for an image. The regions may vary with situations and may include polygons that encompass contiguous pixels, or may be defined by different ranges of intensities which may not even be contiguous. The Cascade Object Detector uses the Viola-Jones [7] algorithm to filter out segments like people's faces, key points on the face like noses, eyes or even the mouth. Viola-Jones object detection framework was the first object detection framework proposed in 2001 by Paul Viola and Michael Jones. It can be modified to detect a diverse range of objects. It was designed primarily to solve the problem of face detection. Viola-Jones algorithm proves to be robust, is real time and is efficient in terms of feature selection and scale and location invariant detection with a very high detection rate.

3.3. Filtering

3.3.1 Introduction to Filtering.

Filtering is a technique used for modifying or enhancing an image. It can be used to filter an image to emphasize certain features or remove other features. Image processing operations implemented with filtering include smoothing, sharpening, and edge enhancement. Many image processing (filtering) operations are modelled as linear systems. Filtering is a neighborhood operation, in which the value of any given pixel in the output image is determined by applying some algorithm to the values of the pixels in the neighborhood of the corresponding input pixel. A pixel's neighbor-hood is some set of pixels, defined by their locations relative to that pixel. Linear filtering is filtering in which the value of an output pixel is a linear combination of the values of the pixels in the input pixel's neighborhood. Certain image processing operations involve processing an image in sections, called blocks or neighborhoods, rather than processing the entire image at once. Several functions such as linear filtering and morphological functions use this approach.

Linear filtering of an image is accomplished through an operation called convolution. Convolution is a neighborhood operation in which each output pixel is the weighted sum of neighboring input pixels. The matrix of weights is called the convolution kernel, also known as the filter. A convolution kernel is a correlation kernel that has been rotated 180 degrees. The operation called correlation is closely related to convolution. In correlation, the value of an output pixel is also computed as a weighted sum of neighboring pixels. The difference is that the matrix of weights, in this case called the correlation kernel, is not rotated during the computation.

Images are often corrupted by random variations in intensity, illumination, or have poor contrast and can’t be used directly. Filtering can be used to transform pixel intensity values to reveal certain image characteristics, to improve contrast and re-move noise. This can result in an enhanced and smoothed image.

Many image processing operations result in changes to the image's histogram. The class of histogram modifications which are considered here include operations where the changes to pixel levels are computed so as to change the histogram in a particular way. Digital images are prone to a variety of types of noise. Noise is the result of errors in the image acquisition process that result in pixel values that do not reflect the true intensities of the real scene. There are several ways that noise can be introduced into an image, depending on how the image is created. For example, if the image is scanned from a photograph made on film, the film grain is a source of noise. Noise can also be the result of damage to the film, or be introduced by the scanner itself. There are many other digital sources for such noise as well. If the image is acquired directly in a digital format, the mechanism for gathering the data (such as a CCD detector) can introduce noise. Electronic transmission of image data can also intro-duce noise. To simulate some of the aforementioned effects and to add various types of noise to an image, the imnoise function can be used. Hence noise is removed by Linear Filtering, Median Filtering and Adaptive Filtering. Some of the other common filters used in the field are Mean Filter, Median Filter and Gaussian Filter. Hence a conclusion can be reached that medfilt2 and rangefilt give best results when used to preprocess images for EOH-SIFT.

3.3.2 Medfilt2.

This function performs two-dimensional median filtering on the image being processed. Median filtering [10] is a nonlinear operation often used in image processing to reduce "salt and pepper" noise. Median filtering is more effective than convolution when the goal is to simultaneously reduce noise and preserve edges. This filtering technique is a noise reduction technique which uses a similar processing pattern as rangefilt. Medfilt gets the median of the 8 neighbors and assigns the current pixel being analyzed, the median value. This effectively removes a lot of the noise in the picture that could have caused inaccuracies.

3.3.3 Rangefilt.

This function computes the local intensity range in a neighborhood around each pixel in an image as in [11]. The neighborhood is defined by the domain binary mask. Elements of the mask with a non-zero value are considered part of the neighborhood. By default a 3x3 matrix containing only non-zero values is used. At the border of the image, extrapolation is used. After processing, the current pixel is set to the difference of the maximum and minimum values of the neighbors. Hence when the brightness of an image changes, the difference between the maximum and minimum values will remain constant. This means, rangefilt will give the same output of a given image irrespective of the brightness or illumination of the picture taken. Hence, EOH-SIFT can be made illumination invariant by making use of rangefilt. Also, the high contrasting output produced by rangefilt means that the accuracy of the EOH-SIFT algorithms will also improve.

$$F(X) = Max(X_i) - Min(X_i) \quad \textbf{(1)}$$

Where, X is the array of pixels and Xi takes values of the 8 neighboring pixels, i.e., the pixels around the current pixel being filtered.

Representation of the effects of Filters.

(b) Unfiltered Image

(a) Filtered Image

Fig. 1. Before (a) and after (b) filtering

.

Therefore, it can be seen in Table 1 that the matching and overall effect of EOH-SIFT has improved appreciably.

3.4. EOH SIFT For Feature Extraction

3.4.1 SIFT.

With respect to image processing, SIFT (Scale Invariant Feature Transform) algorithm is the most widely used algorithm available today. Scale Invariant Feature Transform (SIFT) is useful algorithm which was proposed by David G. Lowe of the University of British Columbia [1]. The features of SIFT are local and are based on important key points which remains unaffected by scale and variations in view point. Scale invariance implies that if a keypoint is generated at a corner, and an enlarged version of the image is again processed, SIFT will not give multiple key-points at that corner. The SIFT algorithm can be segmented into multiple separate processes. Firstly, the image is converted into many vector collections which are not dependent on the rotation, translation, scaling and to some extent to the illumination (as shown in equation 2,3,4 and 5). Edge response points and low contrast points are discarded. The second process is feature matching and indexing and involves finding the nearest best neighbors by using the Best-bin-first search algorithm. Lowe has refined the threshold values of this algorithm so as to eliminate the majority of the false matches. Clusters of key-points are identified by Hough transform voting. The generated model can be verified by using the linear least squares which reduces the sum of the squared distance of locations in the model to the image locations. Outliers are the key-points that are far away from the other key-points clusters. These outliers have a lesser probability of being valid key-points and can be eliminated.

For scale-space extrema detection,

$$D(x,y,\sigma) = L(x,y,k\_i\sigma) - l(x,y,k\_j\sigma) \qquad \textbf{(2)}$$

Where, L(x, y, kσ) is the convolution of the original image I(x, y) with the Gaussian blur G(x, y, kσ) at scale kσ.

$$\mathrm{L(x,y,k}\sigma) = \mathrm{G(x,y,k}\sigma) \times \mathrm{I(x,y)} \qquad \textbf{(3)}$$

To achieve rotation invariance the following are precomputed.

$$\mathrm{m(x,y)}=\sqrt{(L(x+1,y)-L(x-1,y))^2+(L(x,y+1)-L(x,(y-1))^2} \qquad \textbf{(4)}$$

$$\theta\mathrm{(x,y)=atan(2}\times\mathrm{(L(x,y+1)\text{- }L(x,y\text{-}1),L(x+1,y)\text{- }L(x\text{-}1,y)))} \qquad \textbf{(5)}$$

Where, m(x, y) is the gradient magnitude and θ(x, y) is the orientation.

3.4.2 EOH.

The EOH (Edge Oriented Histogram) [12] features rely on gradient information obtained according to the Sobel operator, which creates an image emphasizing the edges. Different orientation bins are fixed, and a rectangular candidate sub window is given with the gradient magnitude of each pixel assigned to an orientation bin which is closest to gradient orientation of the pixel. It becomes both easy and fast to compute with the use of one integral image per orientation bin to compute the gradient magnitude summation for any candidate sub window. Thus different kinds of features can be defined in comparison to Haar filters.

3.4.3 EOH-SIFT.

EOH-SIFT have better results than SIFT, as it assigns a main orientation to key-points. It intuitively assigns every key point to the same orientation. Thus EOH is limited to matching key-points between images that suffer translation misalignment mainly. This algorithm can give varied results in some cases. To increase its accuracy, [5] have come up a technique where they have combined SIFT and EOH. This technique was found to give more accurate results. It has been found that by using various filters to improve the quality of the input image and by applying these algorithms only on the region on interest (a face in our case), a more definitive result is obtained. Hence, the accuracy of SIFT-EOH combination can be improved by doing some pre-processing and filtering.

3.5. Matching

The matching technique used finds nearest key-points from different spectral images in the descriptor space by filtering feature vectors with low descriptive elements. This process is based on the Euclidian distance between the corresponding descriptor vectors. In order to increase the matching robustness, two key-points are matches only if the ratio of the first and second best matches is of a lesser value than a given threshold. Furthermore, the matching robustness is increased by discarding those keypoints that have some of their sub-regions without information (i.e. sub-regions only containing a few contours). SIFT can robustly identify objects even among clutter and under partial occlusion, because the SIFT feature descriptor is invariant to affine distortion and illumination changes.

Key-points are categorized into the following 3 main categories. Flat region - the region in the center of the image where there is no intensity change, Edge region - the edge of an image where there is no change in intensity along the direction of the edge and Corner region - the corner of an image where there is no significant change in intensity in at least 2 directions.

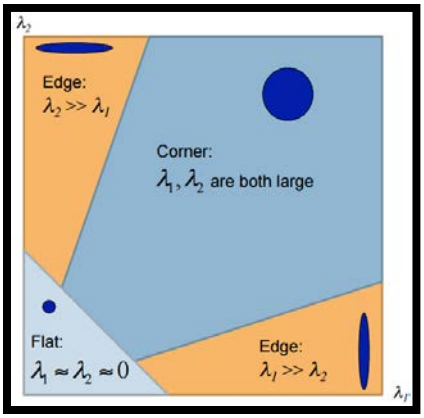


Fig. 2. Categories of key-points

Cornerness Function:

$$R = \lambda_1\lambda_2 - \kappa(\lambda_1 + \lambda_2)^2 \quad \textbf{(6)}$$

Where, κ is a constant between 0.04 and 0.15.

3.6 *Methodology*

The series of steps involved in our approach is depicted in the figures given below.
Pseudo code of Face recognition is explained here.

- Let gr be the set of gallery images.

```
For each gri in gr:
        Rfi ← Rangefilter(gri)
        Mfi ← Medianfilter(Rfi)
        EdgeEmphasis ← EdgeOrientedHistogram(Mfi)
        Keypoints ← SIFTfeaturedetector(edgeEmphasis)
        Descriptor ← SIFTfeaturedetector(Keypoints)
        Match_output ←siftMatch(Descriptor)
End for
```

3.6. Results and Inferences

Table 1. Comparison of the of Filtered EOH-SIFT to EOH-SIFT and SIFT

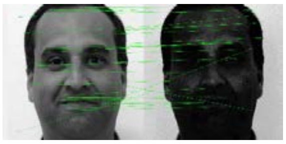
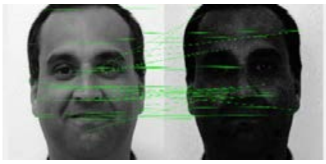
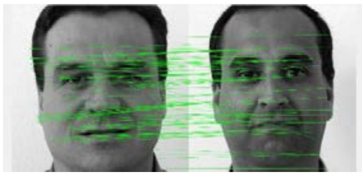
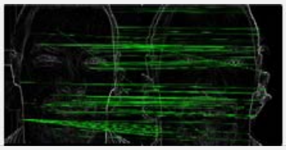
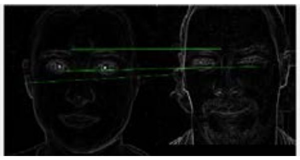
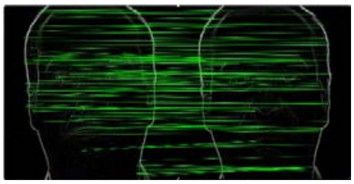
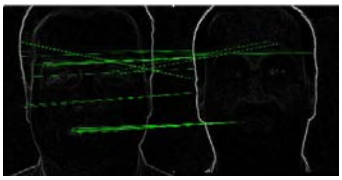
METHOD	MATCHED	MIS-MATCHED
		
SIFT		
		
EOH-SIFT		
		
Filtered EOH-SIFT		

Table 1 shows the significant improvement in the results compared to the non-filtered images. Furthermore the proposed approach was implemented on datasets shown in Table 2 which confirms the improvement in results. The figures just illustrate the results obtained with only two pairs of images using SIFT, EOH and the proposed approach (quantitative evaluation over the whole data set is presented in Table 2). Note that since these pairs correspond to rectified images the matching should correspond to keypoints lying in the same row; in other words the segments that connect keypoints should be horizontal lines.

Table 1. Results

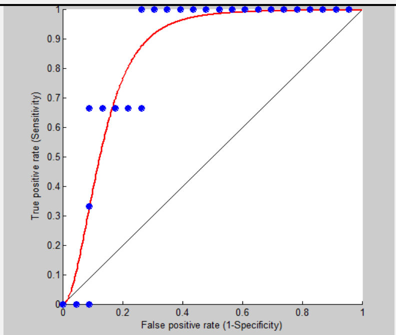
Database	Accuracy	Precision	Recall
FEI Dataset	86.94%	81.81%	89.98%
CALTECH Dataset	82.11%	72.27%	84.98%

The dataset available through the FEI Database was used for evaluation of accuracy, precision and recall of the proposed approach. The performance of the approach has been tabulated in Table 2. By the use of filters it has been seen that even low illumination images with other noises give more positive results compared to non-filtered images.

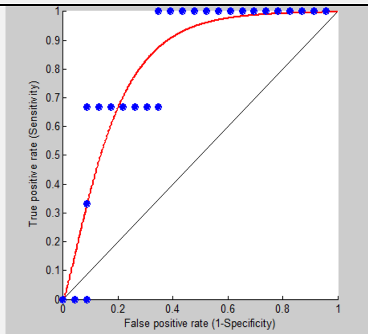
There are two common plots that are often used in face recognition systems namely, Receiver Operating Characteristics (ROC) and Cumulative Match Characteristics (CMC). The ROC curve is used to examine the relation between the sensitivity and the specificity. It is used to describe the quality of a 1:1 matcher. On other hand CMC plot depicts the ranking ability of an 1:m identification system. These are depicted in Table 3 and 4 respectively.

Table 3.ROC Curves

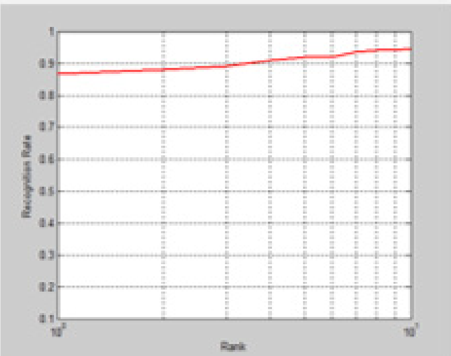
ROC Curve for FEI Dataaset



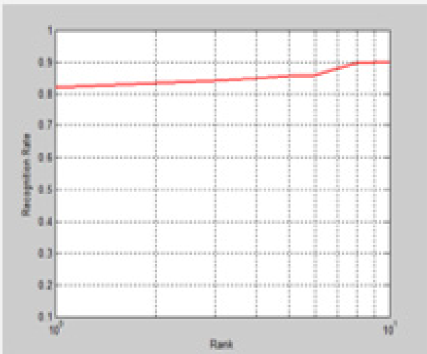
ROC Curve for CALTECH Dataset



CMC Plot for FEI Dataset



CMC Plot for CALTECH Dataset



4. Conclusion and Future Work

The filters described in this paper have proved to be of substantial use when matching faces. In Table 1, it can be observed that more number of horizontal lines (usually correct matches) are obtained when the approach in this paper is used when compared against EOH-SIFT. Hence, the approach described in this paper improves upon the EOH-SIFT approach and is more geared towards handling face recognition. This paper does not include a comparison with other approaches due to various configuration settings that might have been used in that research work. There is much scope of providing a more comparative study in the future.

Future work will be aimed towards using Kernel based Affine Invariant SIFT features with FEOH and also deployment of various filters to understand the effect of these filters for improving efficacy of the system.

5. References

1. Lowe, David G. (1999). "Object recognition from local scale-invariant features", Proceedings of the International Conference on Computer Vision 2. pp. 1150–1157. doi:10.1109/ICCV.1999.790410.
2. Ke, Y., and Sukthankar, R., "PCA-SIFT: A More Distinctive Representation for Local Image Descriptors", Computer Vision and Pattern Recognition, 2004.
3. Abdel-Hakim, A.E.; Farag, A.A., "CSIFT: A SIFT Descriptor with Color Invariant Characteristics." in Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on , vol.2, no., pp.1978-1983, 2006.
4. "Mexican Government Adopts FaceIt Face Recognition Technology to Eliminate Duplicate Voter Registrations in Upcoming Presidential Election." Business Wire. 2000-05-11. Retrieved 2008-06-02.
5. Aguilera, C.; Barrera, F.; Lumbreras, F.; Sappa, A.D.; Toledo, R. "Multispectral image feature points." Sensors 2012, 12, 12661–12672.
6. Mortensen, E.N., Deng, H., Shapiro, L. (2005). "A SIFT descriptor with global context". In Computer Vision and Pattern Recognition (CVPR 2005), 20-25 June 2005. IEEE, Vol. 1, 184-190.
7. Paul Viola, Michael Jones "Robust Real-Time Face Detection" ,International Journal of Computer Vision 57(2), 137–154, 2004.
8. Gangqiang Zhao, Ling Chen, Gencai Chen, Junsong Yuan, "KPB-SIFT: A Compact Local Feature Descriptor", In the Proceedings of the 18th ACM international conference on Multimedia, Pages 1175-1178, 2010.
9. Jean-Michel Morel, Guoshen Yu, "ASIFT: A New Framework For Fully Affine Invariant Image Comparison", in Society for Industrial and Applied Mathematics Journal, Vol. 2, No. 2, pp. 438-469,2009.
10. Gajanand Gupta, "Algorithm for Image Processing Using Improved Median Filter and Comparison of Mean, Median and Improved Median Filter", in International Journal of Soft Computing and Engineering (IJSCE), Volume-1, No.5, ISSN: 2231-2307,2011.
11. D G Bailey, R M Hodgson, "Image and Vision Computing", Volume 3, Issue 3, August 1985, Pages 99-110
12. Levi, K, Weiss, Y, "Learning object detection from a small number of examples: the importance of good features", 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol.2, Page(s):II-53 - II-60 ,ISSN :1063-6919.