Image Similarity Assessment Based On Feature Based Sparse Representation

K. Radha
PG Scholar
Department of Information Technology
K.S.Rangasamy College of Technology
Tiruchengode.

J. Nithya
Associate Professor
Department of Information Technology
K.S.Rangasamy College of Technology
Tiruchengode.

Abstract **: Image similarity assessment is essentially important to several multimedia information processing systems and applications.The major goal of image similarity assessment is to design algorithms for repeated and objective evaluation of similarity in a consistent manner with subjective human evaluation.In the existing method, similarity metrics Human visual system (HVS) and Natural scene statistics (NSS) are efficient to measure the "quality" of an image evaluated with its original version, particularly for some image restoration applications. The HVS and NSS mainly focus on evaluating the similarities between a reference image and its non-geometrically variational versions, such as decompressed and brightness/contrast-enhanced versions. In several applications, assessment of the similarities between a reference image and its geometrically variational versions, such as conversion, turning round, sizing, spinning, and other deformations, is required. A feature based sparse representation(FBSR) approach is proposed to measure the information present in an image, based on robust image quality extraction. The identification of the feature points in an image is followed by describing each feature point using a descriptor. The descriptors of an image is represented via sparse representation and the similarity is assessed between the two images using sparse coding technique. The feature based approach mainly focus on reducing the computational complexities for lexicon feature extraction and image matching by performing sparse coding, which can be further reduced using new algorithm such as the Efficient greedy algorithm and Online lexicon learning algorithm.**

Key Terms **: SIFT Feature Extraction , Lexicon Learning , Image Similarity Assessment , Sparse Representation , Image copy Detection .**

1.INTRODUCTION

In the domain of image processing, assessment of image similarity is significant for many multimedia applications. The similarity measurement is to automatically assess the similarities among the images in a consistent manner. A simple and popularly used metric is the peak signal-to-noise ratio (PSNR) or the corresponding mean squared error(MSE).MSE is a standard criteria for assessment of image quality and fidelity measure[1].The goal of a signal fidelity measure is to compare two signals by providing a quantitative score that describes the degree of relationship/ reliability or, on the other hand, the phase of inaccuracy/deformation between images.It is simple, It is parameter free and inexpensive. The MSE possesses the features to satisfying properties of convexity, symmetry, and differentiability. The MSE is also a desirable measure in the

statistics and estimation framework. It has been employed extensively for Optimizing and assessing a wide variety of signal processing applications. The relationship between MSE/PSNR and human judgement of quality is not suitable for most applications[1]. Image Quality Assessment (QA) algorithms are used for understanding the similarity with a reference or perfect image[2]. Full-Reference QA methods attempt to achieve consistency in quality prediction by modelling salient physiological and psychovisual features of the human visual system (HVS), or by signal fidelity measures[2].

To Quantify the loss of image information to the distortion process, QA systems are regularly involved with judging the visual quality of 'natural' images and videos for 'human utilization. The Human Visual System(HVS)/Natural scene statistics (NSS) mainly focus on assessing the similarity between a reference image and its non-geometrically variational versions to improve the PSNR metric[3]. Some advanced approaches, Structural Similarity Index (SSIM) and Visual information fidelity(VIF) are used to capture the loss of image structure.

The VIF method is better than a HVS based method and VIF performs well in single-distortion as well as in cross distortion scenarios. SSIM/VIF can tolerate slightly geometric variations[2].The feature based sparse representation(FBSR) approach is proposed to extract the information that is present in a reference image and how much of this information can be extracted from the test image to measure the similarity between the two images.The FBSR adopt the Scale Inariant Feature Transform(SIFT)[4].The SIFT features are invariant to image size and turning round, and to afford strong matching across a sizeable range of affine alteration, change in 3D viewpoint, noise, and change illumination. The features are highly unique and a single feature can be correctly matched with high probability against a large database of features from many images. SIFT features are widely used in several multimedia applications, such as image retrieval, recognition, copy detection, and near-duplicate detection[4]. The image is characterized by a set of perspective invariant region descriptors so that recognition can proceed successfully even though changes in viewpoint illumination. The similarity with text retrieval implementation matches on descriptors are pre-computed and the scheme builds method of indexing descriptors extracted from local regions, and is robust to background clutter[5]. The local region descriptors are hierarchically quantized in a term tree. The term tree allows a larger and more inequitable words to be used efficiently, it leads to a dramatic improvement in

retrieval quality and the tree is directly defines the quantization. The quantization and the indexing are consequently fully integrated[6].

Overview of Proposed Scheme

In several Multimedia applications, the assessment of similarity between a reference image and its geometrically variational versions is required. A feature based sparse representation approach is proposed to measure the information present in an image, based on vigorous image quality extraction. The identification of the quality points in an image is followed by describing each quality point using a descriptor. The descriptors of an image is represented by sparse depiction and the similarity is assessed between the two images using sparse coding method. The quality descriptor is sparsely depicted in terms of a lexicon or transferred as a linear combination of lexicon atoms, and to achieve efficient feature representations and vigorous image similarity assessment. Atom means a unique sample or basic unit learned from a set of training data.

A lexicon consisting of several atoms can provide sparse depiction of the data as a linear amalgamation of a few atoms.The feature based approach mainly focus on reducing the computational complexities for lexicon feature extraction and image matching by performing sparse convention, which can be further reduced using new algorithms, such as the Efficient greedy algorithm and Online lexicon learning algorithm.

II PROPOSED QUALITY BASED SPARSE REPRESENTATION ON IMAGE SIMILARITY ASSESSMENT (FBSR) METHODOLOGY:

Sparse representation has resulted in major impact on computer vision and pattern recognition. The goal of FBSR is to obtain the compact representation of the observed signal and also to extract semantic information.The selection of lexicon plays a key task to achieve the image similarity assessment.In the proposed scheme , the sparse representation and lexicon learning techniques are used. FBSR scheme first apply the standard SIFT to detect the keypoints of an image based on the keypoints, the quality points are extracted from the test and reference image. To make the quality points more compact using the lexicon learning algorithm and finally using the sparse coding to achieve the image similarity[3].

Scale Invariant Feature Transform(SIFT)

SIFT transforms image data into scale-invariant coordinates virtual to local quality points and generates large numbers of points that compactly cover the image over the full range of scales and locations[3]. The quantity of features is mostly important for entity identification, where the ability to distinguish small objects in cluttered backgrounds requires that at least 3 points be correctly matched from each entity for reliable identification.

A. *SIFT Feature Extraction*

SIFT features are first extracted from a set of reference images and stored in a database. Fig1 shows the reference image. A new image is harmonized by independently evaluating each feature from the new image to this previous database and finding candidate matching features based on Euclidean distance of their feature vectors.SIFT feature extraction has been performed in four stages [3]

1. *Scale-space extrema detection*

Search over all scales and the entire image locations. Fig-2 Shows the difference-of-Gaussian function to identify potential interest points that are invariant to scale and orientation. The scale space of image is defined as a function of L(u,v,z), that is produced from the convolution of a variable scale Gaussian, G(u,v,z), L(u,v,z)=G(u,v,z)*l(u,v), where * is the convolution operation in u and v and

$$G(u,v,z) = \frac{1}{2\pi z^2} e^{-(u^2+v^2)} \quad \ldots\ldots\ldots\ldots(1)$$

To efficiently detect stable keypoint locations in scale space, In the proposed scheme using scale-space extrema in the difference-of-Guassian function convolved with the image (u,v,z), which can be computed from the difference of two nearby scales separated by a constant multiplicative factor k.

$$C(u,v,z) = (G(u,v,kz) - G(u,v,z)) * I(u,v) = L(u,v,kz) - L(u,v,z) \ldots\ldots\ldots(2)$$

Fig 1. Reference Image

Confined extrema detection

Each sample point is compared to perceive the confined maxima and minima.The samples are compared with eight neighbors of the current image and nine neighbors of the scale above and below. Maxima is selected only if it is larger than all neighbors and minima is selected only if it is smaller than all neighbors.

2 .*Keypoint localization*

A keypoint has been found by comparing a pixel to its neighbors and is to perform a detailed fit to the nearby data for location, scale, and ratio of key curvatures. The low dissimilarity points or poorly restricted behind an boundaries are removed by key point localization.

$$C(U)=C+\frac{\partial D^T}{\partial U}\ U+\frac{1}{2}U^T\frac{\partial^2 D}{\partial^2 U}U\ldots\ldots\ldots......(3)$$

Fig 2. Diffrence of Gaussion

Where C and its imitatives are estimated at the sample point U and $U=(u,v,z)^T$ is the offset from the sample point.

Eliminating Edge response:
The stability is improved by removing the low contrast keypoints.The difference-of-Gaussian function will have a strapping response along boundaries, even if the position along the boundaries are poorly determined and therefore unstable to small amounts of noise.

3*Orientation assignment*

Based on local image slant information, one or more orientations are allocated to each key peak location. All potential actions are performed on image data that has been changed relative to the dispersed orientation, size, and position for each surface, thereby providing invariance to these revolutions.

4.*Keypoint descriptor Generation*

The key points are transformed into a representation that allows for significant levels of local shape distortion and change in illumination.

Representation of Feature Descriptor

A keypoint descriptor is created by first computing the gradient magnitude and direction at each image sample point in a region around the keypoint location. The key points are weighted by a Gaussian window, indicated by the overlaid circle and the samples are accumulated into orientation histograms summarizing the contents over 4x4 subregions with the length of each corresponding sum of the gradient magnitudes near that direction within the region. The 2x2 descriptor array computed from an 8x8 set of samples, the proposed scheme use 4x4 descriptors computed from a 16x16 sample array and each descriptor is 128 dimen sional vector. Usually hundreds to thousands of keypoints may be extracted from an image. .

To make the scale invariant feature transform quality is more compact, the bag of words (BoW) representation approach quantizes SIFT descriptors by vector quantization technique into a collection of visual words based on a pre-defined illustration terms or term tree[4] .

Term Tree

The term tree defines a hierarchical quantization that is built by hierarchical k-means clustering. A large set of consign descriptor vectors are used in the unsupervised training of the tree. The indexing descriptors extracted from local image region and is robust to background clutter.The local region descriptors are hierarchically quantized into a term tree. The term tree allows a larger and more inequitable term to be used efficiently, which leads to a striking enhancement in seizure quality. The most important property of the scheme is that the tree directly defines the quantization.The quantization and the indexing are fully integrated.

4.1.2 BoW representation

BoW representation approach quantizes SIFT feature descriptor by vector quantization into a collection of visual words based on the visual term tree.The BoW representation approach assessing the similarity between SIFT feature descriptors can be measured by matching their corresponding visual words by histogram matching.

Dessriptor Testing

In the descriptor testing , two parameters used to vary the complexity of the descriptor. 1. The number of orientations r in the histograms and the width n of the n $\times$ n array of orientation histograms.2.The size of the resulting descriptor Vector is rn2. The complexity of the descriptor groes, it will be able to discriminate better in a large database and it will also be more sensitive to shape distortions and occlusion[3].

B. *Lexicon Learning*

The feature vectors are represented in terms of a lexicon atoms.The lexicon atoms are Constructed for a set of SIFT features using the K-Singular Value Decomposition lexicon learning algorithm. K-SVD simplify the K-Means algorithm[10].The K-SVD is an iterative process that alternates between sparse coding of the prototypes based on the current lexicon and an update process for the lexicon atoms to better glowing the data. The update of the lexicon columns are joint with an update of the sparse representation coefficients correlated to it, resulting in rushed convergence. The K-SVD algorithm is stretchy and can effort with any detection scheme. Given a set of K SIFT feature vectors v_i and i=1, 2,3…..k and apply K-SVD algorithm to find the lexicon D. K-SVD has been performed in two steps.

1.*Sparse coding stage*

The feature vectors are sparsely decomposed into training samples by using the orthogonal matching pursuit and the samples are grouped into more discriminative samples.

2. *Lexicon update stage*
Update the lexicon for improved robust training samples.The lexicon is together with the non-zero coefficients of sparse representation samples. The size of the lexicon is L×N,L<N<<K, L is the length of the SIFT and N is the Sparse coefficient length, by formulating the problem as

$$D, [u_i]\min_{i=1,2,3\ldots k}\left(\sum_{i=0}^{n}\|v_i - Du_i\|_2^2\right)$$
$$\text{focus to } \forall_i, \|u_i\|_0 \leq L, \ldots\ldots\ldots(4)$$

Where $u_i \in R^{N \times 1}$ is the sparse representation coefficients of v_i and l_0- is the norm of u_i , counts the number of nonzero coefficients of u_i and L is the most desired number of nonzero coefficients of u_i. During the lexicon learning the l_0-minimization in (4) can be converted into an l_1- minimization problem.The obtained lexicon feature D is an overcomplete lexicon where the lexicon contains N Prototype feature vector atoms as the column vectors in D. Each feature vector can be sparsely represented in terms of lexicon atoms D satisfying $\|v_i - Du_i\|_2 \leq$ $\in$, where $\in \geq 0$ is an error tolerance.

ALGORITHM I: PROPOSED FSR

Input: A reference image IM_1 and a test image IM_2 .
Output: The similarity value between IM_1 and IM_2 , i.e., $SIM(IM_1 , IM_2)$
.1. Extract the SIFT feature vectors v_{1i} , i=1,2,3…K_1 , , from IM_1 , followed by learning the lexicon feature sparsely representing .
2. Extract the SIFT feature vectors v_{2i} , i=1,2,3…K_2, , from IM_2, followed by learning the lexicon feature sparsely representing .

Online lexicon learning:

Require: u $\in R^m \sim p(\mathbf{x})$ $\lambda \in R$

(regularization parameter), $D_0 \in R^{m \times k}$ (initial lexicon), T (number of iterations).
1: $A_0 \in R^{k \times k} \leftarrow 0$, $B_0 \in R^{m \times k} \leftarrow 0$ (reset the "past" information).
2: for $t = 1$ to T do
3: Draw u_t from $p(u)$.
4: **Sparse coding:** compute using LARS
$$\propto_t \triangleq \arg\min_{\propto \in R^k} \frac{1}{2}\|u_t - D_{t-1}\|_2^2 + \lambda\|\alpha\|1.$$
5: $A_t = A_{t-1} + \propto_t \propto_t^T$
6: $B_t = B_{t-1} + u_t \propto_t^T$
7: Compute D_t using lexicon update , with D_{t-1} as warm restart, so that
$$D_t \triangleq \arg\min_{D \in C} \frac{1}{t}\sum_{i=1}^{t}\frac{1}{2}\|u_i - D \propto_i\|_2^2 + \lambda \|\propto_i\|_1.$$
8: end for
9: Return D_T (learned lexicon).

Lexicon Update:

Require: D = [d1, . . . ,dk] $\in R^{m \times k}$ (input lexicon),
A = [a1, . . . ,ak] $\in R^{k \times k}$
B = [b1, . . . ,bk] $\in R^{m \times k}$
1: repeat

2: for $j = 1$ to k do
3: renew the j-th column to optimize D_t
$$m_j \leftarrow \frac{1}{2\,A[j,j]}(b_j - Da_j) + d_j (\text{Orthogonal projection}$$
vector)
4: end for
5: until convergence
6: Return D (updated lexicon).
3: Perform l_1-minimization by solving (3) for v_{2j}
with respect to $D_{12} = [D_1, D_2]$
4: Calculate the reconstruction errors, E_{1j} and E_{2j}, for v_{2j}, with respect to D_1 and D_2 , respectively.
5:Perform voting by comparing E_{1j} and E_{2j}, for, v_{2j}, and get the percentages of votes, VT_1 and VT_2, with respect to D_1 and D_2 , respectively.
6:Calculate $SIM(IM_1 , IM_2) = \frac{VT_1 - VT_2 + 1}{2}$ ….(4)

C. Sparse Representation Based Image Similarity Assessment

Consider the two SIFT feature descriptors v_{1i}, i= 1, 2,3…K_1, and v_{2j}, j=1,2,3…K_2 extracted from the reference image(IM_1) and test image(IM_2) with length L. K_1 and K_2 are the number of feature descriptors of test and reference image .The lexicon features of IM_1 and IM_2 are D_1 (of size $L \times N_1$) And D_2 (of size $L \times N_2$) where L< N_1 and L< N_2.Hence $v_{1i} = D_1 u_{1i}$ and $v_{2j} = D_2 u_{2j}$, where u_{1i} and u_{2j} are two sparse coefficient vectors with length N_1 and N_2

Assess the similarity between a reference and test image, use the discriminative features of sparse representation[9]. Concatenating the lexicon atoms for Measure how much of information present in the reference image can be exracted from the test image by greedy algorithm. Concatenating lexicon features for representing each SIFT feature descriptors(v_{2j}) of test image. The joint lexicon $D_{12} = [D_1| D_2]$ can be defined as
$$\hat{u}_j = \text{subject to } \|u_{2j} - D_{12}u_j\|_2$$
$$\leq \in \ldots\ldots(5)$$
Where u_j is the sparse coefficient vector of test image . To solve the sparsest solution for u_i , (2) can be transmit to an l_1-minimization problem as

$$f_n(D) \triangleq \frac{1}{n}\sum_{i=1}^{n} l(u_i D)\ldots\ldots..(6)$$

where **D** in $R^{M \times k}$ is the lexicon, each column representing a basis vector, and is a loss function such that $l(u,D)$ should be small if D is good at representing the signal u in a sparse fashion. The number of samples n is usually large, whereas the signal length m is relatively small, but each signal only uses a few elements of D in its representation and the overcomplete dictionaries with $k > m$ are allowed.define $l(u,D)$ as the optimal value of the l_1 sparse coding problem

$$l(u,D) \triangleq \min_{\propto \in R^k} \frac{1}{2}\|u - D\|_2^2 + \lambda\|\propto\|_1$$

where λ is a regularization parameter problem is also known as basis pursuit and $\propto$ is the sparse decomposition coefficient.The basis pursuit is well known that $\ell 1$
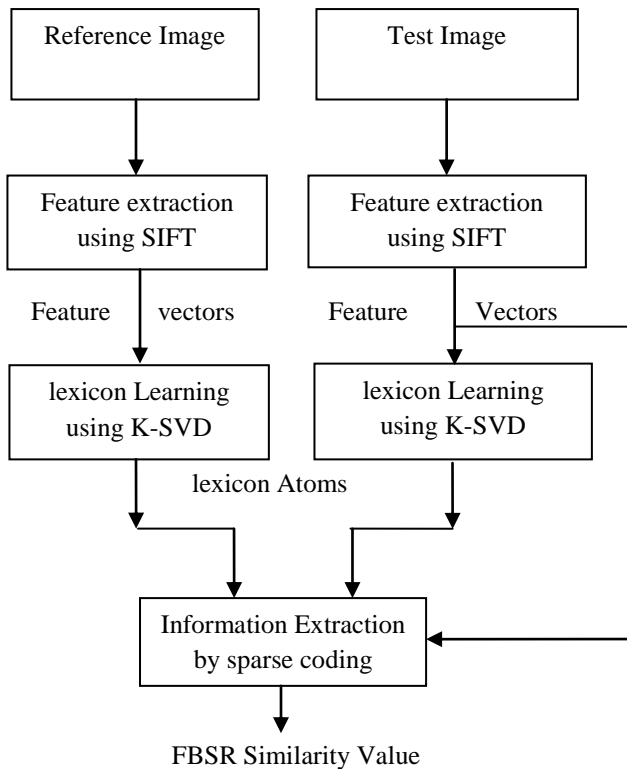
Fig 3.Concept of proposed scheme

regularization yields a sparse solution for α, but there is no straight logical relation between the value of λ and the corresponding effective sparsity ||α||0.To prevent **D** from having arbitrarily large values, Pursuit is common to constrain its columns **d**1, . . . ,**d**k to have an ℓ2-norm less than or equal to one, and call C the convex set of matrices verifying this constraint:

$$C \triangleq \{\mathbf{D} \in R^{m \times k} \text{ s.t. } \forall\, j = 1, \ldots, k,\, d_j^T\, \mathbf{d}j \le 1\}$$

The problem of minimizing the empirical cost $fn(\mathbf{D})$ is not convex with respect to **D** and It can be rewritten as a joint optimization problem with respect to the lexicon **D** and the coefficients $\alpha = [\alpha 1, \ldots, \alpha n]$ in $R^{K \times N}$ of the sparse decompositions, which is not equally curved, but curved with respect to each of the two variables **D** and α when the other one is fixed.

$$\hat{u}_j = argmin_{D \in C,\ \propto \in R}^{K \times n} \sum_{i=0}^{n} (\tfrac{1}{2}\|u_i - D_{12} u_j\|_2^2$$

$$+\text{T}\ \|u_i\|_1) \ldots\ldots(7)$$

Where T is the real parameter and apply efficient greedy algorithm to directly solve (7) in order to find the sparse representation ($\hat{u}_j$) of v_{2j} with respect to the lexicon D_{12}.The position of non-zero coefficients in sparse representation highly concentrated on only one sub lexicon and the remaining coefficients are zero.The sparse representation is perceptive to expect that the atoms for sparsely representing feature vectors of test image should be selected from the sub lexicon of test image. Based on the sparse representation, calculate the reconstruction error as $\|v_{2j} - D_{12}\,\hat{u}_j\|$.
Reconstruction error of reference image is calculated using the lexicon atoms from the reference image. Reconstruction error of test image is calculated using the lexicon atoms from the test image[11]. Reconstruction error of reference image is greater than test image , the reference image atoms are more suitable to representing the feature vectors of test image and reference image lexicon atoms get vote. Reconstruction error of test image is greater than reference image, the test image atoms are more suitable to representing the feature vectors of test image and the test image get vote. Considering all the SIFT feature vectors of test image , the obtained percentage of votes from D_1 and D_2
are denoted by and VT_1 and VT_2 respectively. Based on the voting strategy define the similarity between the two images, IM_1 and IM_2 , as

$$\text{SIM}(IM_1, IM_2) = \frac{VT_1 - VT_2 + 1}{2} \quad \ldots..(8)$$

Achieve better reconstruction errors for lexicon atoms of test image ,the K-SVD using three rules
1.The number of lexicon atoms in reference should be larger than test image.
2. The K-SVD perform number of iterations for learning lexicon atoms of reference image should be larger than test image.
3.More number of non-zero coefficients are used to representing the each feature vector for learning lexicon atoms of reference image.
Considering all the SIFT feature vectors of test image and obtained percentage of votes from test and û image lexicon atoms. The vote for reference and test image range is 0<=1.The similarity range of two images [-1,1] which can be shifted to [0,1].The larger similarity shows that more lexicon atoms are learned from reference image.The result indicates that a considerable amount of information existing in reference image can be extracted from the test image. More lexicon atoms are learned from test image and the result shows that less /no informations are extracted from test image.

III Conclusion

The core proposed a feature -based image similarity assessment technique by exploring the two aspects of a feature detector in terms of representation and matching in FBSR . The reference image features were extracted by using Scale Invariant feature Transform. SIFT features are first extracted from a set of reference images and stored in a database. A new image is matched by individually comparing each feature from the new image to this previous database and finding candidate matching features based on Euclidean distance of their feature vectors.SIFT feature extraction first search over all scales and image locations. A difference-of-Guassian function to identify potential interest points that are invariant to scale and orientation.To efficiently detect stable keypoint locations in scale space.In the proposed scheme using scale Extrema in the difference-of-Guassian function convolved with the image, which can be computed from the difference of two nearby scales separated by a constant multiplicative factor k and the key point has been found by comparing a pixel to its neighbors. BoW representation approach quantizes SIFT feature descriptor by vector quantization into a collection of visual words based on the

visual vocabulary tree.The BoW representation approach assessing the similarity between SIFT feature descriptors can be measured by matching their corresponding visual words by histogram matching. The SIFT is one of the most pervasive and vigorous image features , and it has been widely used in several multimedia applications, image retrieval , appreciation , copy detection and near-duplicate detection.The proposed FBSR may be extended to video copy detection by learning the lexicon features for each video sequence/clip.

REFERENCES

[1] Z.Wang and A. C. Bovik, "Mean squared error: Love it or leave it?—A new look at signal fidelity measures," *IEEE Signal Process. Mag.*, vol. 26, no. 1, pp. 98–117, Jan. 2009.

[2] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430–444, Feb. 2006.

[3] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[4] J. Sivic and A. Zisserman, "Video Google: A text retrieval approach to object matching in videos," in *Proc. IEEE Int. Conf. Computer Vision*, Nice, France, Oct. 2003, vol. 2, pp. 1470–1477.

[5] D. Nistér and H. Stewénius, "Scalable recognition with a vocabulary tree," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2006, pp. 2161–2168.

[6] L. W. Kang, C. Y. Hsu, H. W. Chen, and C. S. Lu, "Secure SIFTbased sparse representation for image copy detection and recognition,"in *Proc. IEEE Int. Conf. Multimedia and Expo*, Singapore, Jul. 2010.

[7] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 10, pp. 1615–1630, Oct. 2005.

[9] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. Huang, and S. Yan, "Sparse representation for computer vision and pattern recognition," *Proc. IEEE*, vol. 98, no. 6, pp. 1031–1044, Jun. 2010.

[10] M. Aharon, M. Elad, and A. M. Bruckstein, "The K-SVD: An algorithm for designing of overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322,Nov. 2006.

[11] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.

[12] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online learning for matrix factorization and sparse coding," *J. Mach. Learn. Res.*, vol. 11, pp. 19–60, 2010.

[13] A. Yang, A. Ganesh, Y. Ma, and S. Sastry, "Fast 1 -minimization algorithms and an application in robust face recognition: A review," in *Proc. IEEE Int. Conf. Image Process.*, Hong Kong, Sep. 2010.

[14] C. Y. Hsu, C. S. Lu, and S. C. Pei, "Secure and robust SIFT," in *Proc. ACM Int. Conf. Multimedia*, Beijing, China, 2009, pp. 637–640.

[15] C. Y. Hsu, C. S. Lu, and S. C. Pei, "Homomorphic encryption-based secure SIFT for privacy-preserving feature extraction," in *Proc. IS&T/ SPIE Media Watermarking, Forensics, and Security*, Jan. 2011.