# Oil Portrait Snapshot Classification on Mobile

Yan Sun and Xiaomu Niu
*School of Electronic Engineering and Computer Science, Queen Mary University of London, London, U.K.*

Keywords: Scale-Invariant Feature Transform, Perception Hash, Support Vector Machines, Bag of Features, Perceptual Hash Distance.

Abstract: In recent years, several art museums have developed smartphone applications as the e-guide in museums. However few of them provide the function of instant retrieval and identification for a painting snapshot taken by mobile. Therefore in this work we design and implement an oil portrait classification application on smartphone. The accuracy of recognition suffers greatly by aberration, blur, geometric deformation and shrinking due to the unprofessional quality of snapshots. Low-megapixel phone camera is another factor downgrading the classification performance. Carefully studying the nature of such photos, we adopts the SIPH algorithm (Scale-invariant feature transform based Image Perceptual Hashing)) to extract image features and generate image information digests. Instead of popular conventional Hamming method, we applied an effective method to calculate the perceptual distance. Testing results show that the proposed method conducts satisfying performance on robustness and discriminability in portrait snapshot identification and feature indexing.

## 1 INTRODUCTION

How to effectively analyze art paintings, understand the subjects, recognize the style and identify the corresponding artists have always been a challenge that many computer scientists are eager to resolve since last century. Quite a few research work have been conducted in this area. For example, (Bartolini et al., 2003) introduced and compared various image processing techniques and algorithms which could be applied in fine art researches. While work (Johnson et al., 2008) presented an algorithm for certificating a specific painting's authenticity. Authors in (Nack et al., 2001) proposed an approach that adopts low-level descriptors to represent prototypical style elements, and high-level conceptual descriptors to meet contextual and intuitive demands. Other prior work (Stork, 2009) (Martinez et al., 2002) (Pelagotti et al, 2008) have been carried out to study the similar problems. However those methods mainly target at professional high-resolution data and in the state of theoretical evaluation. Practical applications which could run on smartphones for multi-resolution digitalized paintings are seldom investigated. Resolution, aberration, light changes, and geometric deformation are the main problems which exist in the photos captured by smartphone camera. Moreover, human perceptual and visual factors play important roles in art identification which further complexes the situation.

In this work, we developed a light weight software on smartphone platform which can process and analyze the oil portrait snapshots captured by the mobile phone camera, aiming to classify the artist or style of the paintings with feedback to the mobile phone users. In order to reduce the interference caused by the problems mentioned above, we adopt the SIPH (SIFT (Scale-Invariant Feature Transform) based Image Perception hash) to generate image digests. The extracted features from SIPH are resistant to geometric distortions such as shearing, scaling and rotation, as well as with strong robustness to non-geometric attacks including illumination changes, affine and projection effects. Unlike the method of calculating perceptual distance in image hash algorithm which only employs the global threshold (Phash.org, 2016), we developed a more effective method to calculate the perceptual distance, with about 16% improvement under blur influence, 8% improvement under mosaic and 7% improvement under rotation and shearing as shown in the testing results.

Besides the painting retrieval solution, a learning-based classifier is also desired to classify a portrait



Sun Y. and Niu X.
Oil Portrait Snapshot Classification on Mobile.
DOI: 10.5220/0006082401430149
In *Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2017)*, pages 143-149
ISBN: 978-989-758-225-7

which has no painting records in the current database. Work (Gancarczyk and Sobczyk, 2013) proposed a complex approach based on a data mining algorithm which sets on a pixel level of the image to resolve the tasks of art work identification and restoration. In this paper, we adopted SVMs (Support Vector Machines) classifier in conjunction with Bag of Features (BOF) for machine learning and classification. To our best knowledge, till now no similar work has been presented in the literature. The main contributions of the project are summarized as below:

- We design a light weight client and server structure to allow photos taken by smartphone camera can be processed and analysed at the server side.
- We adopt the SIFT based perceptual hashing method running at the server side to extract prominent characteristics of images shoot by smartphones. The approach has good robustness to geometrical and non-geometrical attacks and can effectively extract and match visual features at different resolutions.
- We apply a new method which takes the human perceptual factors into consideration in the process of perceptual hash distance calculation to further improve the accuracy. The distortion of perceptual distance calculated from our model is analysed by comparing with other methods.
- A prototype implementation of the portrait analysis application has been developed on Android system with C#. The image processing functions are realized at the server side to allow the client side light weighted.

## 2 RELATED WORK

Interest operator is a method to extract significant features of images, which are distinctive among neighborhood pixels. In order to perform image matching properly, the extracted features need to bear a few properties and criteria (Haralick and Shapiro, 1992), such as distinctness, invariance, stability, global uniqueness and interpretability. SIFT (Scale Invariant Feature Transform) operator applies algorithms to detect and describe local features in images by transforming the image into a set of feature descriptors. The algorithm used was firstly published by David Lowe in 1999 (Lowe, 1999), and later improved in 2004. SIFT has good robustness to local geometric distortion as invariant to variable attacks. As a stable and robust interest operator, SIFT has been employed in many different applications (Yun et al., 2007) (Tian et al., 2014) ( Witek et al., 2014)( Susan et al., 2015), yet few for art work analysis. According to the theoretical and experimental results from (Koenderink, 1984) and (Lindeberg, 1998), feature detection using SIFT follows five major filtering approach steps (Hess, 2010). Perceptual hashing (pHash) employs algorithms to compare hashes in order to measure the similarity between two objects. The essential characteristics of pHash algorithm include perceptual robustness, undirectional and collision resistance. pHash can achieve image identification and authentication by recognition and matching of image digests. The PM function is to calculate the perceptual hash distance between two media objects' hash values:

$$pd_{ij} = PM(h_i, h_j) \tag{1}$$

where $h_i$ and $h_j$ indicate the hash values of two media objects, and the perceptual distance $pd_{ij}$ is obtained using different computing methods, such as Hamming distance, Manhattan Distance, Euclidean distance and so on. Hamming is the most commonly used.

In terms of machine learning, SVMs (Support Vector Machines) are supervised-learning models which applies learning algorithms to conduct data analysis for classification or regression. In recent years SVMs demonstrate the capability in pattern recognition and classification problems (Vapnik, 1995). SVMs is developed from Statistical Learning Theory and has many fine characters, such as robustness, accuracy and effectiveness even with a small training set (Vapnik, 1995), which is particularly suitable for a smartphone app.

Derived from the Bag of Word (BOW), Bag of features (BOF) is an algorithm model widely adopted in computer vision area. Nowadays BOF becomes more popular in the field of image classification. The main idea of the BOF model is to treat images as groups of independent patches, and to sample a significant group of patches and generate visual descriptors for each image. The distribution follows the De Finetti's Theorem (Kerns and Székely, 2006) so the joint probability distribution underlying the data can keep stable for transformation. Inspired by above techniques, we employee the BOF based SIFT combining with SVMs to construct the classifier.

# 3 RECOMMENDATION SYSTEM DESIGN

## 3.1 Hashing based Classifier Design

The whole process of image processing system can be divided in two parts, "retrieval" and "classification", as in Figure 1. A modified hashing algorithm combining with SIFT, called SIPH is adopted at the server side. Figure 2 is the framework of proposed SIPH algorithm. Firstly we select an arbitrary hyperplane that passes through the mass center of feature descriptors' distribution to divide image descriptors. Then we value each individual descriptor as 1 or 0 according to its position to the hyperplane. Totally N hyperplanes will be selected to repeat such steps for each descriptor.

Eventually every selected descriptor will be given an N-degree binary sequence, as an N*l hashes of the image (image "fingerprint"). In general SIFT algorithm, similarity of two images is measured by calculating those interest points' nearest neighbors. And in traditional hash algorithm, rate of similarity between two hashes is measured using the Hamming distance (Choi and Park, 2012). For two tested image hashes, the normalized Hamming distance (HD) between them is defined equation (2):

$$HD = \frac{\sum_{n=1}^{L} H_l(n) - H_m(n)}{L} \tag{2}$$

where $H_l$ and $H_m$ are the hash values of the $l_{th}$ and $m_{th}$ images, $L$ is the length of the hash value. The normalized hamming distance $HD$ between two image hashes is negatively correlated to the similarity rate of two images. The value of HD is approaching to 0 when the two images become nearly identical. Referring to (Watson, 1993), in this work, θ=0.5 is chosen to be the threshold: if the HD < $\theta$ , the two images are considered as similar; on the contrary, if HD > $\theta$ , or they are considered as different.

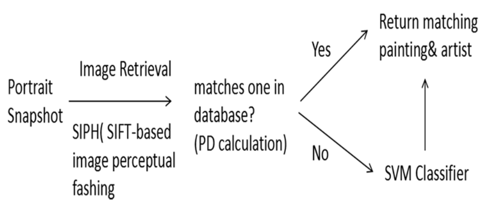


Figure 1: Overall Framework of the System.

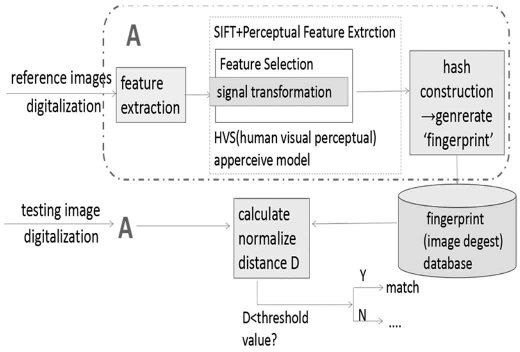


Figure 2: Framework of SIPH algorithm.

In this work, we propose a novel method to calculate the distance for similarity. Given two image hash sequence b and b' generated by SIPH procedure, we consider each generated hash sequence as a number of m-length binary strings. The number of binary strings, the number of feature points and the size of the string of images can be different from each other. For the given sets of binary string of two images, b and b ', we first find each element's corresponding nearest neighbor in the other set (N-length "0/1" string). Nearest neighbors are defined as one element's corresponding binary strings in another collection which have minimum hamming distance to the given element, and the distance is less than a predefined threshold β. Due to the high dimensionality of the hash sequence of an image, the global threshold method is not always the best solution. Therefore we employed a method with more effectiveness with the following steps:

1) Calculate the hamming distance between given element and its nearest neighbour HD1 and hamming distance between given element and its second nearest neighbour HD2.

2) Then compare HD1/HD2 to a certain threshold β1. If HD1/HD2<β1 is met, the element matches its nearest neighbours.

3) After that calculate the ratio of matching elements number N' by min {size (b), size (b')}, and the new perceptual distance was defined as PD' (3):

$$PD' = R\_ele = N' / \min\{size(b), size(b')\} \tag{3}$$

where N' is the total number of nearest elements

4) According to the testing results, we calculated the new threshold β2. When β2<PD'<1, as the value of PD' increases, the perceptual content of 2 images are more similar.

## 3.2 Learning based Classifier Design

For those oil portraits which has no matching image in the current database, we design a machine learning based classifier by applying SVMs (Support Vector Machines) in conjunction with Bag of Features (BOF). For the input data of training process, we transfer the SIPH descriptors of portrait images into visual key words using K-means clustering based on BOF model.

Characterized by repetition, texture or brushwork is recognized as a centralized expression of local features, treated as basic element in this work. These elements are clustered into k groups via k-means clustering algorithm to form a word book. Those groups act as visual vocabularies in the word book. In order to allow the images to be described and represented by the visual vocabularies in the word book, each descriptor is expressed with a nearest texture keyword in Euclidean distance which is to minimize the sum of squared Euclidean distances between the descriptors and their nearest cluster centres. In that way, all the brush/texture descriptors of the images can be replaced by visual keywords. After that, by counting the number of descriptors assigned to a centre, one image can be expressed as a histogram of variable frequencies of texture keywords predefined in the visual vocabulary dictionary (codebook). It is assumed that each image can be represented by a histogram:

$$h_i \in R_+^d (i = 1,2,...,l) \tag{4}$$

where $l$ is the total number of the images.

SVMs model was firstly proposed by (Vapnik, 1995), which is widely adopted in many learning tasks (Joachim, 1998) (Huang et al., 2002) (Roy, 2012). The principle of SVM is employing a max-margin classification hyper-plane. Given $(x_i, y_i)(i = 1,2,...,l)$ while $x_i$ represents the training dataset and $y_i$ is the corresponding label. Now we will solve the optimization problem by maximizing the margin:

$$\min_w \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}\xi_i$$
$$s.t. y_i(w \bullet x_i + b) \geq 1 - \xi_i, (i = 1,...,N) \text{ and } \xi_i \geq 0 \tag{5}$$

where C is the penalty factor, $\xi$ is the slack factor (Huang et al., 2002) (Roy, 2012), and $w$ is the normal value to the linear decision hyperplane. However only a linear decision boundary (a hyperplane parametrized by w) can be learned by this equation. In most of the cases, the classes are not always separable by linear boundaries. So the dual equation is adopted as (6):

(6)

$$\min_a \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l} a_i a_j y_i y_j K(x_i, y_j) - \sum_{i=1}^{l} a_i$$
$$s.t. \sum_{i=1}^{l} a_i y_i = 0 \text{ and } 0 \leq a_i \leq C(i = 1,2,...l) \tag{6}$$

where a is Lagrange multiplier, By solving the equation (6) , a can be obtained. Then we have (7) and decision function (8):

$$w = \sum_{i=1}^{l} a_i y_i x_i \tag{7}$$

$$y_j = \sum_{i=1}^{l} a_i y_i K(x_i, x_j) + b \tag{8}$$

where K is the kernel function and b is a basis value.

Given $x_i, x_j \in R_+^d (i, j = 1,2,3,...,l)$ , the dual formulation allows the use of the kernel trick and the kernel function is defined as:

$$K(x_i, x_j) = <\Phi(x_i) \cdot \Phi(x_j)> \tag{9}$$

where $\Phi : R^d \to F$ represents mapping the vector $x \in R^d$ for input space to a high dimensional feature space F.

A good SVM classifier is with satisfied performance in terms of cohesion and separation. The two terms based on the kernel function are defined in (10) and (11) below:

- Cohesion: similarity within the class

$$w(K) = \sum_{\substack{\phi(x_1)\in C_1 \\ \phi(x_i)\in C_1}} <\phi(x_i), \phi(\overline{x_1})> + \sum_{\substack{\phi(x_2)\in C_2 \\ \phi(x_j)\in C_2}} <\phi(x_j), \phi(\overline{x_2})>$$
$$= \frac{1}{N_1}\sum_{x_i \in C_1} K(x_i, x_i) + \frac{1}{N2}\sum_{x_j \in C_2} K(x_j, x_j) \tag{10}$$

- Separation: dissimilarity between classes

$$b(K) = \sum_{\phi(\overline{x_1})\in C_1, \phi(\overline{x_2})\in C_2} <\phi(\overline{x_1}), \phi(\overline{x_2})>$$
$$= \frac{1}{N_1 N_2}\sum_{x_i \in C_1, x_j \in C_2}\sum_{i,j} K(x_i, x_j) \tag{11}$$

where $\overline{x}_i, i = 1,2$ are the centers of two clusters (class), C1 represent Class 1, C2 represent Class 2, N1 represents the number of the cases in class 1 and N2 represents the number of the case in class 2.

To evaluate the quality of the classifier, we make use of the combination of these two criteria as our classification upper threshold, represented as Evl (lower is better):

$$Evl = w(K) / b(K) \tag{12}$$

# 4 VALIDATION AND TESTING

In this work, the Mono for Android (developed by Xamarin) is adopted to build the application client on Android phone while running the core algorithms at server side. C# language is chosen to write the core algorithms of the application as it runs on the basis of CLR (Common Language Runtime), which make it easy to integrate with variable system and projects written in other languages. The SVMs is implemented in the libsvm package, which provides efficient multi-class Support Vector Classification and Regression results. Portrait paintings from two artists Rembrant and Velasquezi are selected for classifying. In total, we acquired 104 paintings of Rembrant and 52 paintings of Velasquez. Those portrait paintings were downloaded from different online galleries. Therefore the size, resolution and quality of images in database vary a lot. This results in a more robust system to be resistant to various attacks geometrically or non-geometrically.

## 4.1 Validation of Hashing based Classifier

We implement three different methods to verify the robustness and accuracy of image matching of proposed solution.

- SIFT only: use SIFT operators to extract the image feature points and quickly match the closed/similar interest points
- DCT based IPH: use DCT based perceptual hash method to match the same image
- SIFT based IPH: use the proposed method - SIFT perceptual hash to match the image

The perceptual distance PD is calculated as equation (8). Table 2: Perceptual Distance of Three Different Methods under various Attacks.gives the PD value between original portrait and the same image after different kinds of attacks using three matching approaches. The robustness and stability can be estimated by the value of PD. The threshold in the experiment was 0.5. Smaller PD value means better robustness while when PD>0.5, the changes of images are intolerable. From Table 2 and Table 2, the proposed SIFT based perceptual hash method performs better than the other two methods. It is also worth pointing out that among all those attacks, the Gaussian Noise is pretty severe for all three approaches.

Table 1: Image identification accuracy of different algorithms under Gaussian attacks.

| | Accuracy | | |
|---|---|---|---|
| | Hamming Distance (θ = 0.5) | Nearest Neighbour (β1 = 0.8) | Proposed Method (β2 = 0.6) |
| Gaussian Noise (μ = 0.1) | 0.97 | 0.98 | 0.96 |
| Gaussian Noise (μ = 0.5) | 0.83 | 0.87 | 0.90 |
| Gaussian Noise (μ = 0.9) | 0.42 | 0.51 | 0.74 |

Table 2: Perceptual Distance of Three Different Methods under various Attacks.

| Attack | Intensity | PD of SIFT only | PD of DCT based IPH | PD of SIFT based IPH |
|---|---|---|---|---|
| JPEG | 1 (quality factor) | 0.046 | 0.022 | 0.058 |
| Gaussian Noise | 0.3 (average value) | 0.596 | 0.796 | 0.466 |
| Median Filtering | 10 times | 0.233 | 0.041 | 0.01 |
| Blur | 10 times | 0.214 | 0.188 | 0.028 |
| Rotation+ Shearing | 10 degree | 0.233 | 0.36 | 0.311 |
| Mosaic | 10 pixel-window | 0.099 | 0.1 | 0.011 |

## 4.2 Oil Portrait Classification Testing

We test the smartphone application on both server side and client side. The phone camera at the client takes the pictures and sends them to the server via wireless connections. At server side, the images are processed by designed methods as introduced above. The classification result is then feedback to the client via wireless channel. Testing results are shown in Figure 3 to 6. Overall the implemented application is functional well with satisfied classification accuracy.

Table 3: Accurate Rates of Classifying 2 Artists' Portraits with Different γ.

| | Accuracy | | | |
|---|---|---|---|---|
| Classes | γ= 2 | γ= 3 | γ= 4 | γ= 5 |
| Rembrandt | 81.97% | 80.33% | 77.05% | 72.13% |
| Velasquez | 83.02% | 82.22% | 79.02% | 74.41% |
| Average | 82.50% | 81.28% | 78.04% | 76.77% |

## 5 CONCLUSION

In this project, we design and implement an oil portrait snapshot recognition application on smartphone. To achieve successful classification of multi-resolution digitalized image, we proposed an SIFT based pHash to extract the image features and employed a SIFT&BOF based SVM classifier. The results of dimensional testing and experiments, the proposed approach has a strong robustness and stability than others algorithms and achieve an 82.5% accuracy on classifying (identifying).

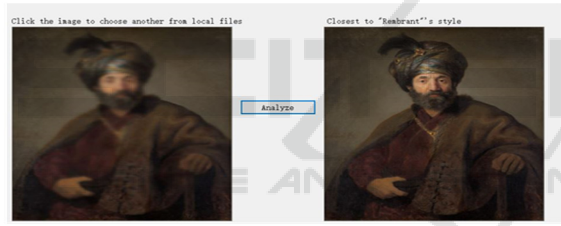


Figure 3: Server side: Noise added Testing Image--find a closed one in training database.

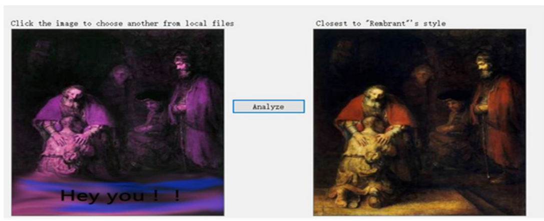


Figure 4: Server Side: Noise and Scaling added Testing Image --find a closed one in training database.

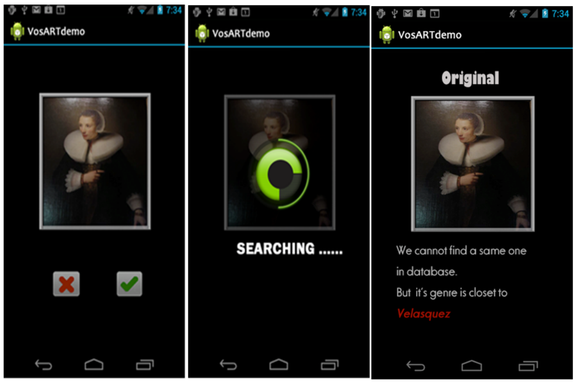


Figure 5: Client Side: Snapshot taken by smartphone -- find one consistent artist.

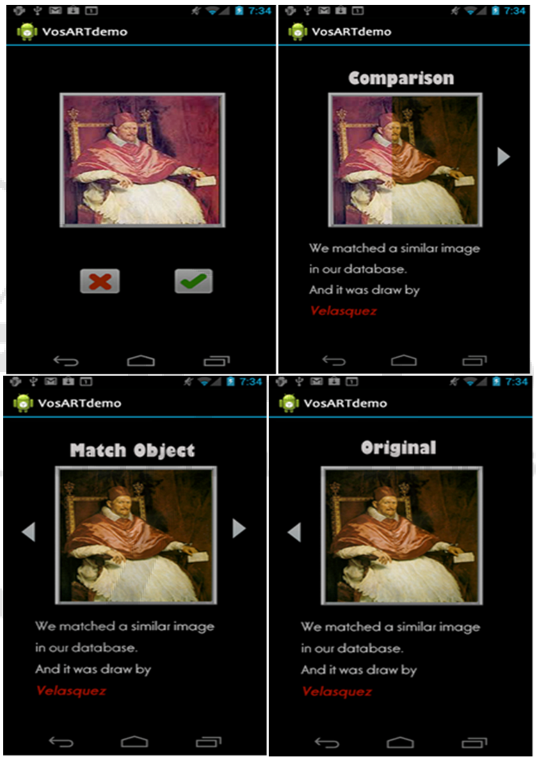


Figure 6: Client Side: Snapshot taken by smartphone -- find a closed one in training database.

## REFERENCES

Bartolini, F. et al., 2003. Applications of image processing technologies to fine arts. In *Optical Metrology* (pp. 12-23). International Society for Optics and Photonics.

Johnson, C. et al., 2008. Image processing for artist identification. *IEEE Signal Processing Magazine*, 25(4), pp.37-48.

Nack, F., et al., 2001. The role of high-level and low-level features in style-based retrieval and generation of multimedia presentations. *New Review of Hypermedia and Multimedia*, 7(1), pp.39-65.

Stork, D. G., 2009. Computer vision and computer graphics analysis of paintings and drawings: An introduction to the literature. In *International Conference on Computer Analysis of Images and Patterns* (pp.9-24). Springer Berlin Heidelberg.

Martinez, K. et al., 2002. Ten years of art imaging research. *Proceedings of the IEEE*, 90(1), pp.28-41.

Pelagotti, A. et al., 2008. Multispectral imaging of paintings. *IEEE Signal Processing Magazine*, 25(4), pp.27-36.

Phash.org., 2014. *pHash.org: Home of pHash, the open source perceptual hash library*. [online] Available at: http://phash.org/ [Accessed 8 Sep. 2016].

Gancarczyk, J., Sobczyk, J., 2013. Data mining approach to Image feature extraction in old painting restoration. *Foundations of Computing and Decision Sciences*, 38(3): pp.159-174.

Haralick, R. and Shapiro, L., 1992. *Computer and robot vision*. 2nd ed. Reading, Mass.: Addison-Wesley Pub. Co., pp.78-120.

Lowe, D. G., 1999. Object recognition from local scale-invariant features. In: *Computer vision, 1999. The proceedings of the seventh IEEE international conference on* (Vol. 2, pp. 1150-1157). Ieee.

Tian, X. et al., 2014. Feature integration of EODH and Color-SIFT: Application to image retrieval based on codebook. *Signal Processing: Image Communication*, 29(4), 530-545.

Yun S.U. et al., 2007. 3D scene reconstruction system with hand-held stereo cameras. In: *Proceedings of 3DTV Conference*, Kos Island, Greece, May 03-07, 2007.

Witek, J. et al., 2014. An application of machine learning methods to structural interaction fingerprints—a case study of kinase inhibitors. *Bioorganic & medicinal chemistry letters*, 24(2), 580-585.

Susan, S. et al., 2015. Fuzzy match index for scale-invariant feature transform (SIFT) features with application to face recognition with weak supervision. *IET Image Processing*, 9(11), 951-958..

Koenderink, J. J., 1984. The structure of images. *Biological cybernetics*, Springer, 50(5), 363-370.

Lindeberg, T., 1998. Feature detection with automatic scale selection. *International journal of computer vision*, 30(2), 79-116..

Hess, R., 2010. An open-source SIFT Library. In *Proceedings of the 18th ACM international conference on Multimedia*, ACM. pp. 1493-1496.

Vapnik, V., 1995. *The Nature of Statistical Learning Theory*. Springer, New York, 2nd edition.

Kerns, G. J., Székely, G. J., 2006. Definetti's Theorem for Abstract Finite Exchangeable Sequences. *Journal of Theoretical Probability*, 19(3): 589-608.

Choi, Y. S., Park, J. H., 2012. Image hash generation method using hierarchical histogram. *Multimedia Tools and Applications*, 61(1): 181-194.

Joachims, T., 1998. Text categorization with support vector machines—learning with many relevant features. In *Proceedings of the 10th European Conference on Machine Learning, Chemnitz*, Berlin Germany. pp. 137–142.

Huang, C., Davis, L. S., Townshed, J. R. G., 2002. An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing*, 23, 725–749.

Roy K., 2012. ART based clustering of bag-of-features for image classification. In *Image and Signal Processing (CISP), 2012 5th International Congress on*. IEEE.

A.B. Watson, 1993. DCT Quantization Matrices Visually Optimized for Individual Images. Proc. SPIE, San Jose, CA, USA, vol. 1913, Jan. 31, 1993, pp. 202–216.