

Design of Binary Robust Independent Elementary Features through Compressive Sensing View

Jinhong Zhang¹, Xueqing Liu², Xiangwei Liu^{1*}

¹ Bo Hai Shipbuilding Vocational College, Huludao, Liaoning, China.

² School of Electronic Information and Electrical Engineering, Shanghai Jiaotong University, Shanghai, China.

* Corresponding author. Tel.: 13167050652; email: lxqlxq21@gmail.com

Manuscript submitted September 9, 2014; accepted January 12, 2015.

doi: 10.17706/ijapm.2015.5.1.67-75

Abstract: Binary Robust Independent Elementary Feature (BRIEF) is designed to be a very simple and efficient feature point descriptor for image. In this paper, compressive sensing theory is used in proposing the reason why it works and guiding the parameter determination of BRIEF. It is proved in this paper BRIEF is the Binarization of Compressive Sensing Sampling. Also, with theoretical basis for BRIEF, there is guidance for us to design new descriptors.

Key words: Compressive sensing, binary robust independent elementary feature (BRIEF), descriptor.

1. Introduction

Local feature descriptors are widely used in pattern recognition and movement tracking in computer vision field. There are all kinds of local feature descriptors. Haar-like Feature [1] is very famous for its good performance in face recognition. Following it, Scale-invariant feature transform (SIFT) [2] is now a cutting edge. Amazed by its fantastic performance and robustness against scaling and rotation, for a long time it is believed the complexity of the SIFT is a necessary for a good performance. So on top of SIFT, people developed Speeded up robust features (SURF) [3] and PCA-SIFT [4]. BRIEF [5] was a shock to the field of local feature descriptors, as it is such a simple descriptor which can provide incredible good performance. But yet there is no explanation for it. In [6], the author called BRIEF "LSH on Kendall's Tau" and by imitating the design of BRIEF proposed a new descriptor called Locally Uniform Comparison Image Descriptor (LUCID).

Compressive Sensing is a very hot area of research recently. The sparsity consideration provided by this theory are widely used in sparse sensing, denoising and real-time tracking [7].

Being such a breakthrough in the field, however, BRIEF is never well explained by any theory why it is efficient, and the theoretical basis provided in [6] is very complex and not straight forward and can't explain the efficiency. In this paper, a connection between this smartly designed local feature descriptor and the theory of compressive sensing which is very solid in theoretical basis is made. And another perspective of how to design a binary local descriptor is proposed.

2. BRIEF is Binarization of Compressive Sensing Sampling

2.1. BRIEF: Binary Robust Independent Elementary Features

BRIEF [5] is a quite simple local feature descriptor. First of all, a test τ on an $S \times S$ patch p is defined as

$$\tau(p; x, y) := \begin{cases} 1 & \text{if } p(x) < p(y) \\ 0 & \text{otherwise} \end{cases} \quad (2.1.1)$$

where $p(x)$ is the pixel intensity in a smoothed version of p at $x = (u, v)^T$. Choosing a set of $n_d(x, y)$ - location pairs uniquely defines as a set of binary tests. BRIEF descriptor is defined to be the $n_d(x, y)$ - dimensional bit string

$$f_{n_d}(p) := \sum_{1 \leq i \leq n_d} 2^{i-1} \tau(p; x_i, y_i) \quad (2.1.2)$$

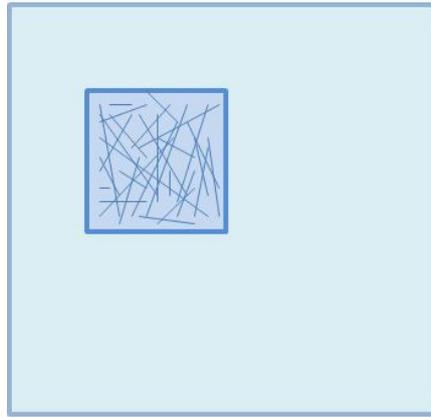


Fig. 1. Every bit of BRIEF is formed by taking pairs of points from a patch randomly and compare their pixel intensity.

As shown in Fig.1, every bit of BRIEF is formed by taking pairs of points from a patch randomly and compare their pixel intensity. If the first point is of smaller pixel intensity, then the bit is 1, otherwise, it is 0.

2.2. Compressive Sensing and Measurement Matrix

Compressive Sensing: Compressive Sensing [8] is a brand new method for data acquiring and sampling. According to it, if the signal is sparse or sparse after transformation, the signal can be project on to a low dimensional space with a measurement matrix while the projection of the signal keeps all the information of the original signal. By solving an optimization problem, the original signal can be reconstructed from the few projections.

Consider a signal x of length n and sparsity $k(k < n)$ projected by a $m \times n(m \ll n)$ measurement matrix A to be

$$y = Ax + \varepsilon \quad (2.2.1)$$

where $\varepsilon \in R^{m \times 1}$ is the error vector, $y \in R^{m \times 1}$ is the vector after sampling. If matrix A follows RIP rules, then x can be reconstructed from y . RIP is defined as below:

If an $m \times n$ matrix A follows

$$(1 - \delta_k) \|z\|_2^2 \leq \|A_T z\|_2^2 \leq (1 + \delta_k) \|z\|_2^2, \forall z \in R^{|T|} \quad (2.2.2)$$

$A \in RIP(k, \delta_k)$, in which according to[9]the sufficient condition for reconstructing x is $\delta_k < 0.307$.

The reconstruction of x is equal to solving a l1 minimization problem,

$$\hat{x} = \operatorname{argmin} \|x\|_1, \text{ s.t. } \|y - Ax\|_2 \leq \varepsilon \quad (2.2.3)$$

Measurement Matrix is vital for Compressive Sensing, there have been a lot of researches on it. In order for compressive sensing to work, the measurement matrix chosen must follow RIP rules in (2.2.2). Up to now, there are several commonly used measurement matrix:

- 1) Gaussian random matrix: Gaussian random matrix is an $M \times N$ matrix A whose every single element follows a 0 mean Gaussian distribution with variance $1/M$.

$$A_{i,j} \sim N(0, 1/M) \quad (2.2.4)$$

And when $M \geq cK \log(N/K)$, Gaussian random matrix follows RIP by great possibility. c is a very small positive number according to [10].

- 2) Bernoulli random matrix is defined similar to Gaussian random matrix except that every element of A follows Bernoulli distribution:

$$A_{i,j} = \begin{cases} +\frac{1}{\sqrt{M}} & P = \frac{1}{2} \\ -\frac{1}{\sqrt{M}} & P = \frac{1}{2} \end{cases} = \frac{1}{\sqrt{M}} \begin{cases} +1 & P = \frac{1}{2} \\ -1 & P = \frac{1}{2} \end{cases} \quad (2.2.5)$$

Or

$$A_{i,j} = \begin{cases} +\sqrt{\frac{3}{M}} & P = \frac{1}{6} \\ 0 & P = \frac{2}{3} \\ -\sqrt{\frac{3}{M}} & P = \frac{1}{6} \end{cases} = \sqrt{\frac{3}{M}} \begin{cases} +1 & P = \frac{1}{6} \\ 0 & P = \frac{2}{3} \\ -1 & P = \frac{1}{6} \end{cases} \quad (2.2.6)$$

Also, when $M \geq cK \log(N/K)$, Bernoulli random matrix follows RIP by great possibility.

- 3) Part orthogonal matrix can also be used as measurement matrix. Part Fourier matrix is a special Part orthogonal matrix. Part orthogonal matrix is developed in the following way:

- Form a $N \times N$ orthogonal matrix U .
- Pick M rows of U to form a $M \times M$ matrix V .
- Normalize V to get part orthogonal matrix A .
- When the size of A is fixed, sparsity K should follow:

$$K \leq c \frac{1}{\mu^2} \frac{M}{(\log(N))^6} \quad (2.2.7)$$

where $\mu = \sqrt{M} \max_{i,j} |U_{i,j}|$, when $\mu = 1$, it is a Part Fourier matrix.

- 4) Part Hadamard matrix can also be used as measurement matrix. It's form in the following way:

- Form a $N \times N$ Hadamard matrix U .
- Pick M rows from U to form Part Hadamard matrix A .

Hadamard matrix can work better than other matrix with the same M on reconstruction. But for Hadamard matrix, N is limited to $2^k, k = 1, 2, 3, \dots$

- 5) Toeplitz matrix and Cyclic Matrix are in the following form:

$$T = \begin{pmatrix} t_n & t_{n-1} & \cdots & t_1 \\ t_{n+1} & t_n & \cdots & t_2 \\ \vdots & \vdots & \ddots & \vdots \\ t_{-1} & t_{2n-2} & \cdots & t_n \end{pmatrix}, \quad C = \begin{pmatrix} t_n & t_{n-1} & \cdots & t_1 \\ t_1 & t_n & \cdots & t_2 \\ \vdots & \vdots & \ddots & \vdots \\ t_{n-1} & t_{n-2} & \cdots & t_n \end{pmatrix} \quad (2.2.8)$$

where T is a Toeplitz matrix and C is a cyclic matrix. They are formed as follows:

- Form a random vector $u = (u_1, u_2, \dots, u_N) \in R^N$
 - Cycle the vector $M(M < N)$ times to form the rest $M - 1$ rows.
 - Normalize through columns to form measurement matrix A .
- 6) Sparse random matrix is formed as follows:
- Form a $M \times N(M < N)$ zero matrix A
 - For each column(or row) of A , choose $d(d < M)$ positions and change the elements to non - zero elements.

2.3. BRIEF Is Binarization of Compressive Sensing Sampling

Since first introduced, people are puzzled why feature descriptor as simple as BRIEF can do such good job. In [6], BRIEF is proved by the author to be "LSH on Kendall's Tau". It is a nice explanation how it is designed, but still, it can't show why it works so effectively. Here we propose another much simpler explanation of BRIEF.

$$y = Ax + \varepsilon \tag{2.3.1}$$

According to the definition of BRIEF, every bit of BRIEF comes from comparing the pixel intensity of a pair of pixels taken from a patch of size $S \times S$. It is the same as line up the $S \times S$ patch to a vector x and sample it with a measure matrix A and then binarize it. The measurement matrix A should be with $S \times S$ columns and M rows where M is the number of bits in BRIEF. Also, the measurement matrix should have a -1 and a +1 in each row.

That makes the matrix a Sparse random matrix with a fixed $d=2$ for each row with the largest element 1 and the smallest element -1.

In [11], a way to find the lower bound for M in Sparse random matrix where all the non - zero elements are positive is calculated to be:

$$M \geq \frac{na^2}{b^2D^2}, (l \leq K) \tag{2.3.2}$$

where a is the smallest non-zero element and b is the largest non-zero element in A , l is the number of non-zero elements in each row, and K is the sparsity of the image. Here we prove this equation also works when there are negative elements in sparse random matrix.

Another version of Restricted Isometry Property is: Assume A to be an $M \times N$ matrix. If for all $x \in R^n$ with less or equal to K non-zero elements, there is a constant $c > 0$ so that

$$c \|Ax\|_2 \leq \|x\|_2 \leq cD \|Ax\|_2 \tag{2.3.3}$$

We call A follows $RIP(K, D)$ where c is a normalization factor and $D > 1$.

It is proved in[11]that if A follows $RIP(K, D)$, then

$$c^2 \|A_j x\|_2^2 \leq 1 \leq c^2 D^2 \|A_j x\|_2^2 \tag{2.3.4}$$

for all $j(1 \leq j \leq n)$ where A_j stands for the j_{th} column of A .

Because of (2.3.4), suppose in each column there are at least $t(1 \leq t < n)$ elements, since the largest absolute value of non-zero elements is a and the smallest absolute value of non-zero elements is b , we

have $ta^2 \leq \|A_j\|_2^2 \leq tb^2$. So $cta^2 \leq 1 \leq ctD^2b^2$ which leads to

$$\frac{1}{c^2} \leq tb^2D^2 \tag{2.3.5}$$

Suppose $s = \min(l, K)$, and suppose the s non-zero elements in x are in the position of the s non-zero elements in A . To make that happen, we can simply shift A so that $A_{1,j} \neq 0 (1 \leq j \leq l)$ and let $x = \sum_{j=1}^s e_j$, then

$$\|Ax\|_2^2 = \sum_{i=1}^m \left(\sum_{j=1}^s A_i^j \right)^2 \geq \left(\sum_{j=1}^s A_i^j \right)^2 \geq (as)^2 \tag{2.3.6}$$

From (2.3.3) and (2.3.5), we have

$$\|Ax\|_2^2 \leq \frac{\|x\|_2^2}{c^2} = \frac{s}{c^2} \leq stb^2D^2 \tag{2.3.7}$$

From (2.3.6) and (2.3.7), we have $(as)^2 \leq stb^2D^2$, in other words

$$t \geq \frac{sa^2}{b^2D^2} \tag{2.3.8}$$

Suppose there are g non-zero elements in A , since there are t non-zero elements in each column and l non-zero elements in each row. Thus $tn \leq g = lm$, in other words

$$t \leq \frac{lm}{n} \tag{2.3.9}$$

From the equations above, we have $\frac{sa^2}{b^2D^2} \leq \frac{lm}{n}$ which leads to

$$m \geq \frac{n sa^2}{l b^2 D^2} \tag{2.3.10}$$

Thus, when $l \leq K, s = l$, so $m \geq \frac{na^2}{b^2D^2}$.

For BRIEF, the measurement matrix has $a^2 = b^2 = 1$ and $l = 2$, thus $m \geq \frac{n}{D^2}$. Since $D > 1$, the lower boundary of m should be smaller than n . Ref. [12] provides a method to estimate K :

- Perform PCA on image patches, find the coefficients and calculate $\sigma = \sqrt{\sum_{i=0}^{n-1} i^2 \alpha_i / \sum_{i=0}^{n-1} \alpha_i}$
- Calculate the sparsity with $K = 0.4066\sigma + 15.6792$

It is quite obvious that the good performance of BRIEF comes from the high reconstruction performance of compressive sensing. Since BRIEF follows the rules for compressive sensing sampling procedure, BRIEF can keep almost all the information to reconstruct the patch, so it is no longer a surprise it can be used as a feature descriptor for pattern recognition or movement tracking and can do a good job. This will be illustrated in the experiment.

3. Experiment and Results

In this part, the performance of BRIEF is shown here as illustrated above.

3.1. Data Set

Here we use the data set used to test BRIEF in [5].

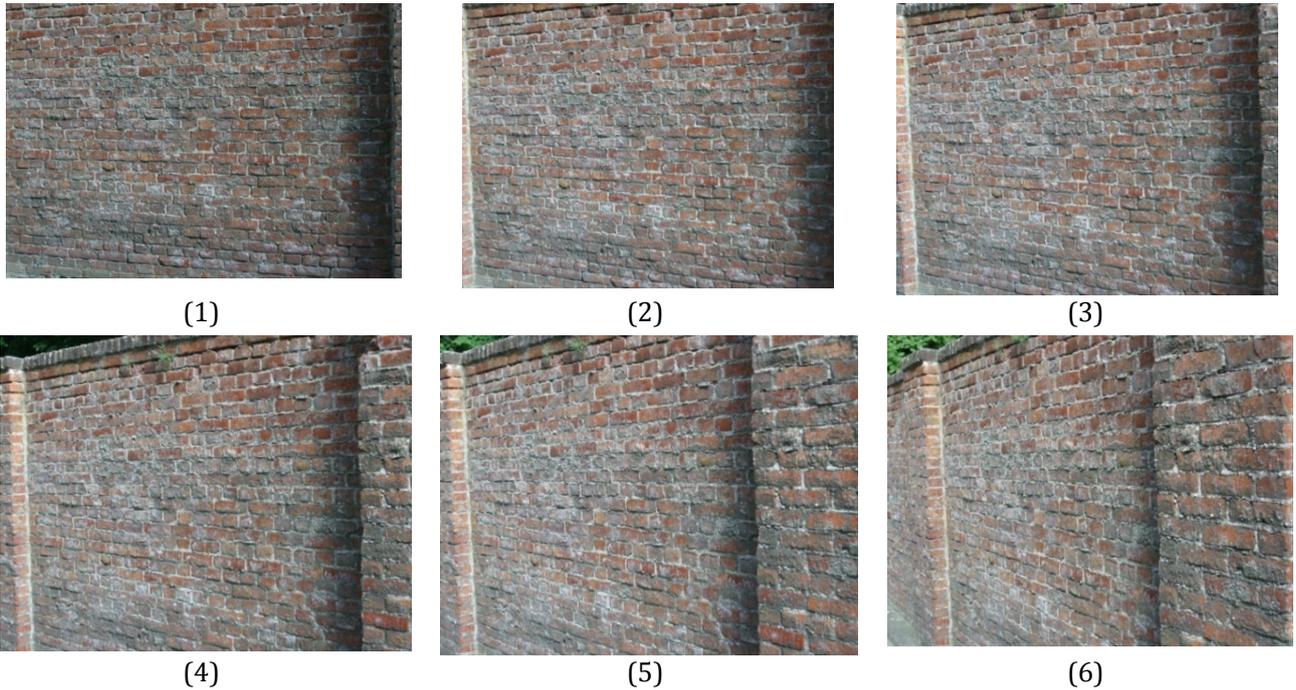


Fig. 2. A group of images from the data set, (1) is the origin image, the others are images with different perspective changes.

3.2. Experiment and Parameters

The patch size here is 32×32 . As the measurement matrix for BRIEF has 2 non-zero elements on each row (a -1 and a +1), we compare it to other binary feature descriptor with measurement matrix which are similar to that but with more elements on each row. We tests these binary feature descriptors with different number of bits.

The FAST feature points as in Fig. 3 are extracted from the first image of the group, then with the homographs provided along with the data set, we transform the feature points into position in other images in the group. Take Fig. 2(2) as an example, by taking 512 of the feature points from Fig. 2(1) and find the 512 corresponding points in Fig. 2(2), the descriptor is calculated for each point. Then by comparing the distances between a point from Fig. 2(1) and all the points from Fig. 2(2), we can find the point in Fig. 2(2) linked by the descriptor to the point from Fig. 2(1). Then we count the percentage of the link that are correct. That's the result reflects the performance of the descriptor. This method is also used in [5].



Fig. 3. Example of the Fast feature points.

3.3. Results

With the PCA methods, we are able to calculate $\sigma = \sqrt{\sum_{i=0}^{n-1} i^2 \alpha_i / \sum_{i=0}^{n-1} \alpha_i} = 8.3798$, thus $K = 0.4066\sigma + 15.6792 = 19.0864$.

In this part, the result of the experiment is shown. For each point in the result, 5 experiments are run respectively, and the result is their average. From Fig. 4 we can see that for the same measurement matrix, the more bits offers better results. Also it is quite clear that when the measurement matrix is sparse enough, l has no effectson M , which fits the theory in (2.3.2).

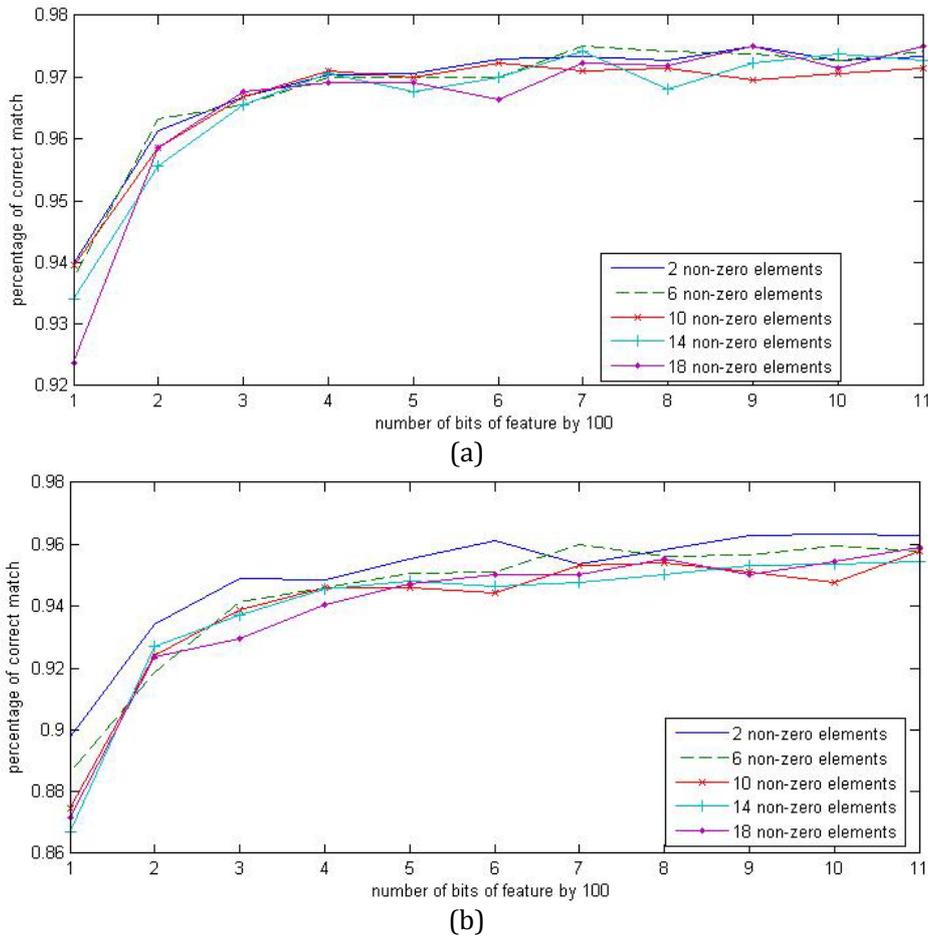


Fig. 4. Experiment results: Here we show (a) results between Fig. 2(1) and Fig. 2(2), (b) results between Fig. 2(1) and Fig. 2(3).

The specific results are shown in Table 1.

Table 1. Performance of the Feature Descriptor

Image NO.	Non-zero ele. in each row	NO. of bits of feature										
		100	200	300	400	500	600	700	800	900	1000	1100
2	2	0.9398	0.9613	0.9668	0.9703	0.9707	0.973	0.9734	0.9727	0.975	0.9727	0.9734
	6	0.9375	0.9633	0.9656	0.9699	0.9699	0.9699	0.975	0.9742	0.9738	0.9727	0.9742
	10	0.9395	0.9586	0.9668	0.9711	0.9699	0.9723	0.9711	0.9715	0.9695	0.9707	0.9715
	14	0.934	0.9555	0.9656	0.9707	0.9676	0.9699	0.9742	0.968	0.9723	0.9738	0.9727
	18	0.9238	0.9586	0.9676	0.9691	0.9691	0.9664	0.9723	0.9719	0.975	0.9715	0.975

3	2	0.8977	0.934	0.9488	0.9484	0.9551	0.9613	0.9535	0.9582	0.9629	0.9633	0.9629
	6	0.8863	0.9184	0.9414	0.9461	0.9508	0.9512	0.9598	0.9563	0.9566	0.9594	0.9578
	10	0.8746	0.9242	0.9387	0.9461	0.9461	0.9445	0.9531	0.9539	0.9512	0.9477	0.9578
	14	0.8668	0.927	0.9371	0.9457	0.948	0.9465	0.9477	0.9504	0.9531	0.9535	0.9543
	18	0.8715	0.9234	0.9293	0.9406	0.9473	0.9504	0.9504	0.9551	0.9504	0.9543	0.959
4	2	0.657	0.7516	0.7922	0.8047	0.809	0.8137	0.8352	0.8277	0.8316	0.8266	0.8324
	6	0.6574	0.7602	0.7742	0.7898	0.802	0.8098	0.8187	0.8141	0.8195	0.8301	0.8207
	10	0.6387	0.732	0.7727	0.7879	0.7941	0.8078	0.8184	0.8176	0.8227	0.823	0.8234
	14	0.632	0.7355	0.7668	0.7879	0.8035	0.7996	0.8105	0.8129	0.8156	0.8246	0.8266
	18	0.6328	0.7406	0.7813	0.7852	0.7996	0.8109	0.8137	0.8187	0.827	0.8195	0.8301
5	2	0.3617	0.4352	0.466	0.4816	0.5008	0.5078	0.5098	0.507	0.5121	0.5148	0.5152
	6	0.3426	0.4277	0.4527	0.4641	0.4816	0.4777	0.4871	0.5012	0.4984	0.493	0.502
	10	0.3461	0.4207	0.4461	0.4617	0.4719	0.4828	0.4926	0.4887	0.4902	0.4984	0.4941
	14	0.3305	0.4023	0.4391	0.4668	0.4715	0.4863	0.4848	0.4879	0.4922	0.502	0.4957
	18	0.3262	0.423	0.4465	0.4762	0.4871	0.4863	0.4895	0.4969	0.4996	0.5047	0.5008
6	2	0.0863	0.1137	0.1246	0.1234	0.127	0.1391	0.143	0.1504	0.1461	0.15	0.1512
	6	0.0969	0.1195	0.1391	0.1355	0.1496	0.141	0.1414	0.1418	0.1492	0.143	0.1496
	10	0.082	0.1113	0.1363	0.1367	0.1445	0.1473	0.1449	0.1453	0.1496	0.1414	0.1512
	14	0.0898	0.1187	0.1371	0.134	0.1398	0.1328	0.1426	0.1449	0.1449	0.1535	0.1535
	18	0.0848	0.1113	0.1344	0.1461	0.1371	0.1527	0.1488	0.1531	0.1527	0.1559	0.1449

4. Conclusion

In this paper, the nature of BRIEF is analyzed from the view of Compressive Sensing. We proved BRIEF is the Binarization of Compressive Sensing Sampling. From the Compressive Sensing point of view, it is quite straight forward to understand why BRIEF's performance is so good. We also did experiment to test the theory. And with the theory we developed above, whenever in the future, there comes a new more effective measurement matrix for compressive sensing, it is possible for us to come up with a new binary feature descriptor which we can use like BRIEF into pattern recognition and movement tracking tasks.

References

- [1] Lienhart, R., & Maydt, J. (2002). An extended set of haar-like features for rapid object detection. *International Conference on Image Processing*, (pp. I-900-I-903).
- [2] Lindeberg, T. (2012). Scale invariant feature transform. *Scholarpedia*, 7, 10491.
- [3] Bay, H., Tuytelaars, T., & Van, G. L. (2006). Surf: Speeded up robust features. *Computer Vision–ECCV 2006*, 404-417. Springer.
- [4] Ke, Y., & Sukthankar, R. (2004). PCA-SIFT: A more distinctive representation for local image descriptors. *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, (pp II-506-II-513).
- [5] Calonder, M., Lepetit, V., Strecha, C., & Fua, P. (2010). Brief: Binary robust independent elementary features. *Computer Vision–ECCV 2010*, 778-792. Springer.
- [6] Ziegler, A., Christiansen, E., Kriegman, D., & Belongie, S. J. (2012). Locally uniform comparison image descriptor. *Advances in Neural Information Processing Systems*, 1-9.
- [7] Zhang, K., Zhang, L., & Yang, M.-H. (2012). Real-time compressive tracking. *Computer Vision–ECCV 2012*, 864-877. Springer.
- [8] Donoho, D. L. (2006). Compressed sensing. *IEEE Transactions on Information Theory*, 52, 1289-1306.
- [9] Cai, T. T., Wang, L., & Xu, G. (2010). New bounds for restricted isometry constants. *IEEE Transactions on Information Theory*, 56, 4388-4394.

- [10] Baraniuk, R., Davenport, M., DeVore, R., & Wakin, M. (2008). A simple proof of the restricted isometry property for random matrices. *Constructive Approximation*, 28, 253-263.
- [11] Jing-Ming, S., Shu, W., & Yan, D. (2012). Lower bounds on the number of observation of sparse random matrices(in chinese). *Signal Processing*, 28, 1156-1163.
- [12] Ma, Y., *et al.* (2013). Image sparsity evaluation based on principle component analysis(in chinese). *Acta Physica Sinica*, 62, 204202.



Jinhong Zhang received her B.S. degree in math, from Bohai University, Jinzhou, PR China. She received her M.S. degree in education, Liaoning Normal University.

She is now an associate professor of Bohai Shipbuilding Vocational College. Her current research interest is math.



Xueqing Liu received her B.S. degree in information and communication engineering from Zhejiang University, Hangzhou, PR China in 2012.

She is now a dual-master student of Georgia Institute of Technology, Georgia, USA and Shanghai Jiao Tong University, Shanghai, China. Her current research interests include computer vision, image and video processing and medical image processing.



Xiangwei Liu received his B.S. degree in mechanical engineering, from University of Science and Technology Liaoning (USTL), Anshan, PR China. He received his M.S. degree from Harbin Engineering University.

He is now an associate professor and an engineer of Bohai Shipbuilding Vocational College. His current research interest is machinery manufacture.