Intelligence Cloud-Based Image Recognition Service

Wei-shuo Li(✉) and Jung-yang Kao

Information and Communications Research Laboratories,
Industrial Technology Research Institute, ITRI, 195, Sec. 4, Chung Hsing Road,
Chutung, Hsinchu, Taiwan, R.O.C.
{ansonli,Yang_Kao}@itri.org.tw

Abstract. Cloud-based vision service provide a opportunity of intelligence and programming support to meet different needs of embedded applications. To reduce the complexity of cloud-based computation, we proposed a method can be by performing Hamming distance. This approach relates in general to a method for feature description, in which a feature patch is described by using a binary string. Our method can achieve near-optimal precision and reduce the bandwidth and computation time.

Keywords: Cloud service · Image recognition · Feature descriptor

1 Cloud-Based Vision Service

The emergence of widespread mobile devices has created vast new opportunities for intelligent applications. When consider the on-device recognition process, for example, the mobile device has the ability to recognize 1,000 images without connecting to the Internet. However, there are applications that need to recognize more than million images. The solution for this is the cloud-based image recognition service [1,2]. All the recognitions will be done in the cloud. Cloud recognition service allows mobile device to work with million of target images stored in the cloud, and provides for a high accuracy recognition rate and very quick response characteristics. This make it very usable to build a conveniently interactive application. Therefore, cloud-based vision can accomplish this with a small memory, near-real time, and low power consumption embedded device [3].

Intelligent vision has been widely used in various application fields of image processing. In general, these applications include a basic process, that is, to extract the features of each image and further compares the extracted feature with a reference feature of the database to locate the best matching target [4]. However, when a large quantity of features is extracted from the images, the required comparison time will be greatly increased. Besides, if the features carry a large volume of data, more bandwidth will be required for transmitting relevant feature description. Therefore, it has become one of the prominent tasks for

J.-H. Lee and S. Pack (Eds.): QShine 2016, LNICST 199, pp. 473–477, 2017.
DOI: 10.1007/978-3-319-60717-7_47

the industries to provide a method for feature description and a feature descriptor using the same capable of expediting feature comparison and reducing the required data volume.

Our key contribution is proposing a descriptor that transforms the input feature space into binary signature such that Hamming distance in the resulting space is closely with similarity measures.

2 Similarity-Preserve Transformation

For clarity, we use SIFT feature as our example, however, our method is applicable to a variety feature descriptor [5,6]. When a keypoint has been detected, we then need describe it by a patch that collecting the nearby pixels. For example, the SIFT algorithm includes (1) take a 16×16 window around detected interest point, (2) divide into a 4×4 grid of cells (3) compute 8 bins angle histogram in each cell. Thus, a SIFT descriptor form a 128 dimension histogram, each bin has 8 bits, and is of 1024 bits size.

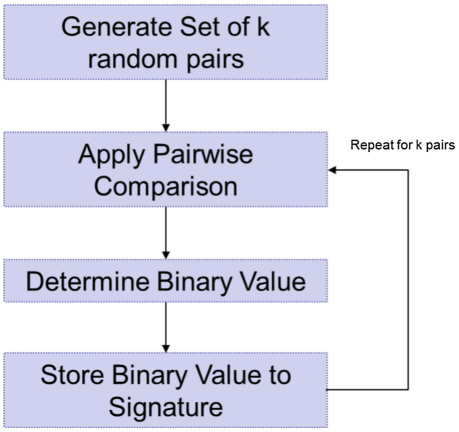

Fig. 1. Flow chart

SIFT need compute the cosine similarity of two histograms to decide a pair of descriptor is similar or not. Our main idea is two similar signatures will preserve similar pairwise relations, thus any descriptor can be represented by binary string of pairwise comparison. From this viewpoint, the original similarity can be approximated by Hamming distance. The key problem is how to choice an appropriate set of pairwise relations. For SIFT, there are $\binom{128}{2}$ pairwise relations, and this information obviously too large. We propose a random projection method, that is, randomly project the original feature space ($[0,1]^d$) to a k-dimension $\{0,1\}^k$ space, and we expect $k < d$, see Fig. 1. Now, the question remains how to determine the value of k to approximately preserve the similarity.

2.1 The Comparative Reason

We shall first consider the question of the probability that random k-projection can preserve similarity. Note that original space, for SIFT, is of size $2^{1024} \gg \binom{128}{2}$, but the feature point is generally sparse, that is, far small than $\binom{128}{2}$ key points(image are typically describe by thousands of keypoints). Thus, we can reasonably assume the pairwise relations can fully distinct two different patches. First, we analysis the probability that similarity is unchanged of a random k-projection method. Denote by S the event fo two descriptors successful matching, and $\delta_H(x, y)$ the Hamming distance. We match a pair of descriptors (x, y) if their hamming distance $\leq h$ with signature size n.

Let k the number of bits select uniformly and randomly from n pairwise relations.

$$\begin{aligned} Pr(S) &= 1 - Pr(\delta_H(x, y) > h) \\ &= 1 - \left(\frac{\binom{n-h}{k-h}}{\binom{n}{k}} \right)^2 \\ &\geq 1 - \left(\frac{\binom{n-h}{k}}{\binom{n}{k}} \right)^2, \end{aligned} \tag{1}$$

using $h = cn, c < 1/2$, and $\binom{n}{k} \leq \left(\frac{en}{k}\right)^k$, the probability $Pr(S)$ is given by

$$\begin{aligned} Pr(S) &\geq 1 - \left(1 - \frac{h}{n}\right)^{2k} \\ &\geq 1 - e^{-\frac{2hk}{n}} \\ &= 1 - \frac{1}{n} \to 1, \text{ if } k = \frac{1}{c} \log n^2. \end{aligned} \tag{2}$$

Equation 2 tell us that we need select larger k, i.e., more bits, as h is deceasing. Since the event S is a tail event, the probability admit a threshold phenomenon, see [7] for more detail. To be more precisely, the behavior of k can be obtained by estimating $Pr(\delta_H(x, y) = h)$, and we can derive additional information of k around the threshold.

We know

$$Pr(\delta_H(x, y) = h) = 0 \text{ for } h > n - k, \tag{3}$$

and

$$\begin{aligned} Pr(\delta_H(x, y) = h) &= Pr(\delta_H(x, y) > h) - Pr(\delta_H(x, y) > h + 1) \\ &= \left(\frac{\binom{n-h}{k}}{\binom{n}{k}} \right)^2 - \left(\frac{\binom{n-h-1}{k}}{\binom{n}{k}} \right)^2 \\ &= \frac{\binom{n-h}{k}}{\binom{n}{k}} \cdot \left(1 + \frac{k}{n-h}\right) \cdot \frac{\binom{n-h}{k}}{\binom{n}{k}} \cdot \left(1 - \frac{k}{n-h}\right) \end{aligned}$$

$$= \frac{\binom{n-h}{k}}{\binom{n}{k}} \cdot \left(1 - \frac{k^2}{(n-h)^2}\right)$$
$$= \left(1 - \frac{k^2}{(n-h)^2}\right) \prod_{i=0}^{k-1} \left(1 - \frac{h}{n-i}\right). \quad (4)$$

This function tell us about the influence of k in the distribution around the threshold.

Table 1. Precision of the random projection method.

# of bits	Precision
128	100%
64	100%
32	93%
16	86%
14	50%
13	30%
12	21%
8	14%

For SIFT 128-bins histogram, choose $k = \frac{1}{c} \cdot \log \binom{128}{2}^2 \approx 32$ for carefully selected c. In practice, $k = 32$ is enough to tackle the recognition problems. In Table 1 we choose 1000 images and run SIFT description process. Then we randomly select difference k compared bits of the 128 bins of SIFT histogram.

Our method is faster than original descriptor. The random projection method only need to perform 32-bits XOR, and is suitable for hardware design. In the case of SIFT, our method is bandwidth-efficient than original descriptor, and we can reduce original SIFT bandwidth to 1/32.

3 Conclusions

The main contribution of our paper is a similarity-preserve transform that transforms the input feature space into binary signature. This paper is directed to a method for cloud-based feature matching service and a feature descriptor using a binary string to describe a feature patch obtained by a feature extraction algorithm. The generated binary string may be used to expedite feature comparison to a near-real time cloud manner. Moreover, since the binary string only requires a small amount of data volume, the required bandwidth may be greatly reduced.

We have shown that (1) our method is similarity-preserve (2)the bandwidth performs comparably to the original SIFT descriptors, (3) computing power degrades gracefully as the number of patch is increased. Moreover, our approach is applicable to a variety of feature description methods.

References

1. Buyya, R., et al.: Cloud computing and emerging IT platforms: vision, hype, and reality for delivering computing as the 5th utility. Future Gener. Comput. Syst. **25**, 599–616 (2009)
2. Foster, I.: Globus online: accelerating and democratizing science through cloud-based services. IEEE Internet Comput. **3**, 70–73 (2011)
3. Soyata, T., et al.: Cloud-vision: real-time face recognition using a mobile-cloudlet-cloud acceleration architecture. In: IEEE Symposium on Computers and Communications (ISCC). IEEE (2012)
4. Bhat, D., Nayar, S.: Ordinal measures for visual correspondence. In: Computer Vision and Pattern Recognition, 2 (1996)
5. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. **60**, 91–110 (2004)
6. Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: Speeded-up robust features (surf). Comput. Vis. Image Underst. **110**(3), 346–359 (2008)
7. Bollobás, B.: Random Graph. Academic Press, London (1985)