

Histogram 에 기반한 Image Hash 개선

An Improved Histogram-Based Image Hash

김 소 영*, 김 형 중

(So-Young Kim and Hyoung-Joong Kim)

Abstract : Image Hash specifies as a descriptor that can be used to measure similarity in images. Among all image Hash methods, histogram based image Hash has robustness to common noise-like operation and various geometric except histogram equalization. In this paper an improved histogram based Image Hash that is using "Imadjust" filter together is proposed. This paper has achieved a satisfactory performance level on histogram equalization as well as geometric deformation.

Keywords: Image Hash, Image Filtering, Information Security, Image Search

I. 서론

image hash는 오랫동안 연구자들에게 흥미있는 주제가 되어왔으며 많은 연구[1,2,3,5,7]들이 있어왔다. Image hash는 content 검색, image 검색, 정보은닉과 image 포렌식등 다양한 분야에서 인증(authentication)을 위한 방법으로 사용된다. 특히 histogram-based image hash는 기하학적 변형에 대한 image histogram 모양의 불변성을 이용함으로써 강력한 image hash 알고리즘[6]으로 제안 되어왔다. 이 방법은 hash의 강인함(robustness)를 향상시키기 위해 선행 단계로 Gaussian filter를 적용하였으며 hash를 생성시 인접하는 두 bin 그룹 사이의 pixel수에서의 관계를 비교하였다. 그러나 이 방법은 histogram equalization에 의한 변형에는 상대적으로 약한 결과를 나타냈다. 본 논문에서는 hash 생성전 선행단계로 Imadjust filter를 적용하여 hash값 생성을 제안한다. 본 논문의 구성은 다음과 같다. II장에서는 기존의 histogram-based image hash 생성방법과 제안된 방법을 살펴보고 III장에서는 제안한 방법의 세부 내용, IV, V장에서는 실험결과를 기존 논문[6]과 비교하며 VI에서 결론을 통해 본 논문을 맺는다.

II. Histogram 에 기반한 Image Hash 생성절차

Image hash 생성은 먼저 선행단계를 거친 영상으로부터 특징을 추출하고 추출된 특징을 사용하여 hash값을 생성하는 것이다. 선행단계는 변형에 대해 hash값이 강인함(robustness)를 갖을 수 있도록 하는 역할을 한다. 다음은 hash 생성 절차를 설명한다.

1. 영상에 필터 적용

Hash의 강인성(robustness)를 보장하기 위한 방법으로 영상에 Low-pass filtering인 Gaussian filtering을 적용한다.

2. 히스토그램 구하기

필터링 작업이 끝나면 필터링이 적용된 영상으로부터 히스토그램을 구할 수 있다. 흑백영상의 픽셀이 0 부터 255의 값을 가지므로 히스토그램은 256 bin을 갖게 되고 된다. Hash 값을 생성하기 위해 256 bin을 L개로 그룹화 한다. 예를 들어 이웃하는 bin을 4개씩 그룹화하고자 하면 L은 '64'가 되고 8개씩 그룹화 하고자한다면 L은 '32'가 된다.

3. Hash값 구하기

두 개의 다른 bin그룹을 비교함으로써 Hash 값을 얻을 수 있다. i번째 bin에서의 픽셀 수를 $h(i)$ 라고 하자. 조건 $1 \leq i < L$ 과 $i < j \leq L$ 을 만족하는 서로 다른 bin을 $\{h(i), h(j)\}$ 과 같이 그룹화 해보자. 이 경우 그룹의 수는 $C_L^2 = L(L-1)/2$ 이 된다. 그룹 $\{h(i), h(j)\}$ 에 대해 $h(i)$ 과 $h(j)$ 을 다음 방법을 통해 hash bit 값을 나타낼 수 있다

$$\text{bit} = \begin{cases} 1, & \text{if } h(i)/h(j) \geq 1 \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

4. Hamming distance 계산

Hash 값을 일차원 벡터값 $\text{hash}(k)$ 라 하면 ($1 \leq k \leq C_L^2$) 한쌍의 hash에 대한 이들간의 Hamming distance값은 다음과 같이 정의할 수 있다.

$$d(\text{hash}_1, \text{hash}_2) = \frac{1}{C_L^2} \sum_{k=1}^{C_L^2} |\text{hash}_1(k) - \text{hash}_2(k)| \quad (2)$$

이때 hamming distance값이 0에 가까우면 비슷한 영상, 0.5에 가까우면 다른 영상이라고 말할 수 있다.

* 책임저자(Corresponding Author)

논문접수 : 20xx. x. x., 채택확정 : 200x. x. xx.

고주몽, 홍길동 : 고려대학교 정보보호대학원

(ko.joomong@korea.ac.kr, hong@korea.ac.kr)

※ 본 연구는 교육부 학술연구조정사업의 지원을 받아 연구되었음.

III. 강인성을 위한 제안된 방법

1. Gaussian Low-Pass filter

앞서 hash값의 강인성을 증대시키기 위한 선행 방법으로 Gaussian filter 적용을 언급하였다. 이 필터링은 기존 논문에서 적용한 방법이다. 본 논문에서도 이 필터를 사용할 것이다.

2. Imadjust filter

영상에서 화질을 균등하게 하는 Imadjust filter는 넓고 현저하게 어두운 부분에 포함된 white, gray를 강화하는데 유용하다. 이 필터는 영상에서 새로운 명암값을 나타낸다. 본 논문에서 선행단계로 Imadjust filter 필터와 Gaussian filter를 각각 또는 동시에 적용해 볼 것이다.

IV. 실험결과

1. 유일성(Uniqueness) 실험

서로 다른 영상중 랜덤하게 두 영상을 선택할 때 각각의 영상이 서로 다른지를 얼마나 잘 판단할 수 있는지에 대한 실험이 유일성(Uniqueness) 실험이다. 본 실험 결과는 총 225개의 흑백영상에 대해 Gaussian filter와 Imadjust filter를 각각 적용한 결과와 함께 적용한 결과이다.

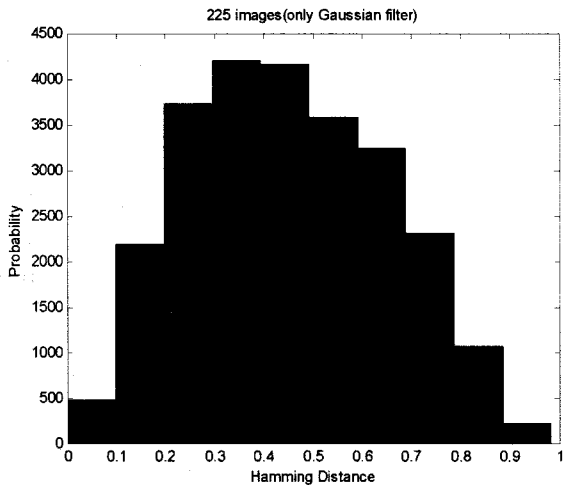


그림 1. 225개 영상에 Gaussian filter만 적용한결과 Hamming Distance 값
Fig 1. Only Gaussian-filtered Different 225images'Hamming Distance

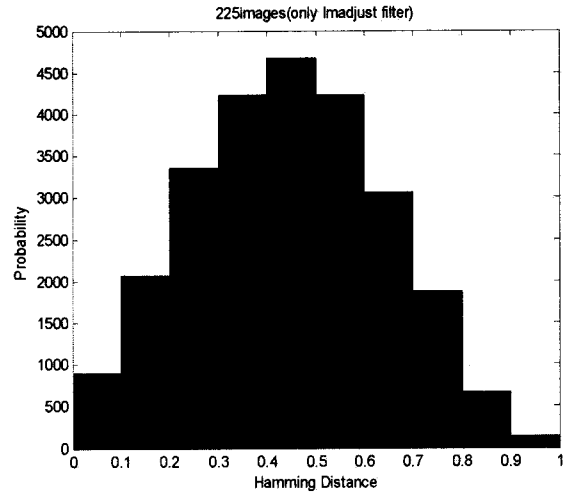


그림 2. 225개 영상에 Imadjust filter만 적용한결과 Hamming Distance 값
Fig 2. Only Imadjust-filtered Different 225images'Hamming Distance

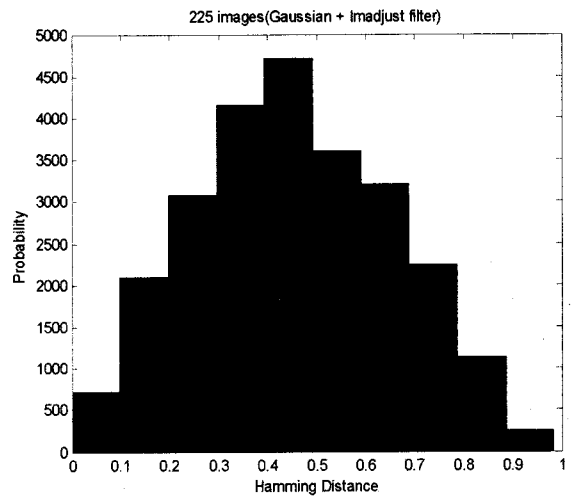


그림 3. 225개 영상에 두 filter 같이 적용한결과 Hamming Distance 값
Fig 3. Gaussian-filter and Imadjust filtered Different 225images'Hamming Distance

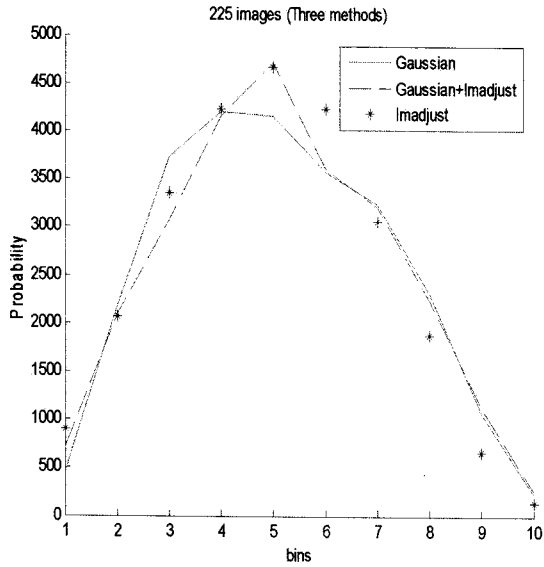


그림 4. 3가지 방법 비교 히스토그램
Fig 4. Comparison of three method's Histogram

서로 다른 영상의 경우 Hamming distance 값이 0.5에 가깝게 나타나야 한다는 [6]의 논문에 의해 히스토그램의 모양이 종상 모양에 가까울수록 좋은 결과라고 할 수 있다.

그림 1,2,3에서 각 방법에 따른 Hamming distance 값을 볼 수 있으며 각각의 히스토그램을 통해서도 차이를 느낄 수 없지만 그림 4를 통해서 두 필터를 함께 사용하는 것이 다른 결과보다 더 좋은 결과를 나타낸 것을 알 수 있다.

2. 강인성(Robustness) 실험

Histogram based image hash 방법은 회전등 기하학적 변형에 강인성을 보이지만 histogram equalization 변형에 대해서는 충분한 스키마가 되지 못했다. 실험은 압축, 필터링, 회전등 변형을 준 일반적인 140개의 복제 영상과 12단계의 Histogram equalization 공격에 대해 별도로 실시하였다. 실험에 사용된 원본영상은 3가지 이미지(Sailboat, Lena, Barbara)이며 원본을 변형하기 위해 사용된 일반적인 방법은 영상1-10(압축률은 달리한 압축), 11-20(quality율을 달리한 압축), 21-40(Gaussian과 Uniform 잡음 적용), 41-70(Gaussian filter, Median filter, With filter 적용), 71-80(회전), 81-90(크기변환), 91-100(shearing), 101-110(crop), 111-140(굴곡, 떨림) 변형을 적용하였으며

Histogram equalization 공격의 12 단계는 (2,5,10,15,30,50,70,100,150,170,200)을 적용하였다.

기존 image hash와 제안된 image hash를 이용하여 원본영상과 변형이 적용된 각각의 140개 영상과 Histogram equalization 공격에 대한 hamming distance를 구한 결과는 다음과 같다.

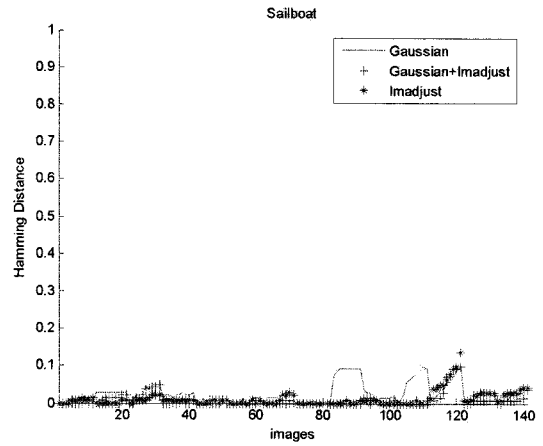


그림 5. Sailboat영상140공격에대한 3가지방법의 Hamming Distance 비교
Fig 5. Comparison of three methods: respectively applying Gaussian filter,Imadjust,and Gaussian,Imadjust together in Sailboat image

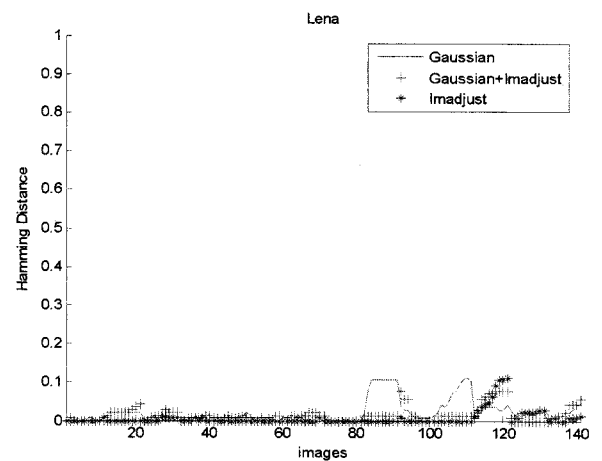


그림 6. Lena영상 140공격 대한 3가지 방법의 Hamming Distance 비교
Fig 6. Comparison of three methods: respectively applying Gaussian filter,Imadjust,and Gaussian,Imadjust together in Lena image

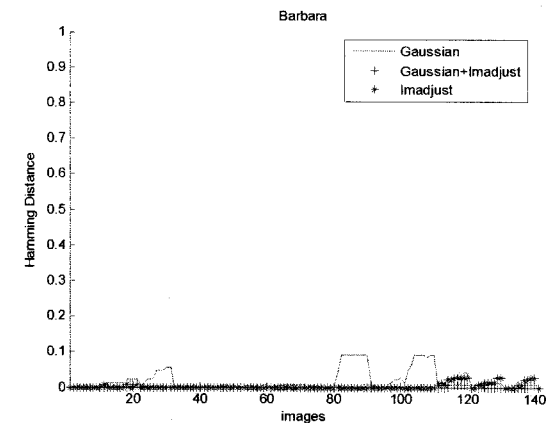


그림 7. Barbara영상 140공격 대한 3가지 방법의 Hamming Distance 비교
Fig 7. Comparison of three methods: respectively applying Gaussian filter,Imadjust,and Gaussian,Imadjust together in Barbara image

그림 5,6,7에서 볼 수 있듯이 선행단계에서 두가지 필터를

같이 적용한 제안된 방법이 기존방법보다 hamming distance 값이 0에 가까이 분포함으로써 공격받은 영상이 원본 영상과 같다는 결과를 더 정확하게 보여준다고 할 수 있다.

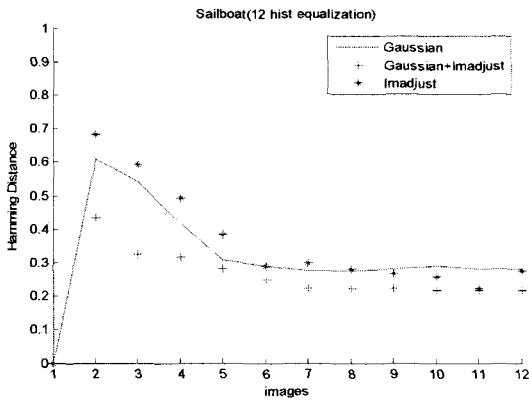


그림 8 Sailboat영상 histogram equalization 대한 3가지 방법의 Hamming Distance 비교

Fig 8. Comparison of three methods: respectively applying Gaussian filter,Imadjust,and Gaussian,Imadjust together in Sailboat image attacked histogram equalization

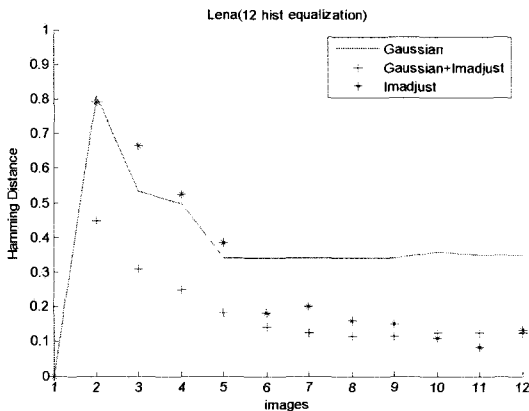


그림 9 Lena영상 histogram equalization 대한 3가지 방법의 Hamming Distance 비교

Fig 9. Comparison of three methods: respectively applying Gaussian filter,Imadjust,and Gaussian,Imadjust together in Lena image attacked histogram equalization

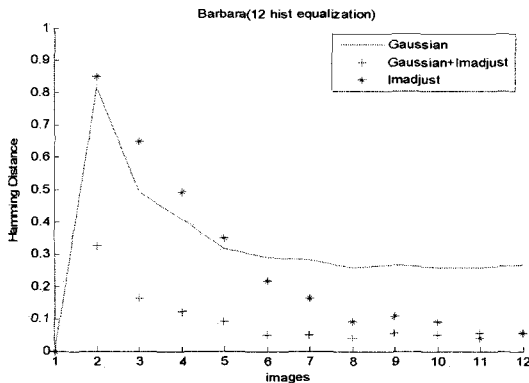


그림 10 Barbara영상 histogram equalization 대한 3가지 방법의 Hamming Distance 비교

Fig 10. Comparison of three methods: respectively applying Gaussian filter,Imadjust,and Gaussian,Imadjust together in Barbara image attacked histogram equalization

그림 8,9,10에서 볼 수 있듯이 12가지의 다양한 Histogram equalization 공격에 있어서도 두가지 필터를 같이 적용한 제안된 방법이 기존방법보다 hamming distance 값이 세영상 모두에서 0.4 이내로 0.9이내로 컸던 기존의 결과보다 좋은 결과를 보여준다고 할 수 있다.

V. 결론

본 논문에서는 매우 간단하고 쉬운 방법으로 개선된 Histogram based image hash를 제안하였다. Hash를 생성하는 선행단계에서 Gaussian filter과 Imadjust filter를 함께 적용함으로써 개 bin의 관계를 이용함으로써 기존 Gaussian filter만을 적용했을때 보다 만족스러운 결과를 얻었다. 특히 제안된 방법으로 Histogram equalization 변형에 취약했던 기존 방법이 상당부분 보완되었다.

참고문헌

- [1] E. Y. Chang, C. Li, J. Z. Wang, P. Mork, and G. Wiederhold . "Searching near replicas of images via clustering." In Proceedings of SPIE Multimedia Storage and Archiving Systems VI, pp. 281-292, Boston, USA, September 1999
- [2] Y. Ke, R. Sukthankar, and L. Huston. "An efficient partsbased ear-duplicate and sub-image retrieval system." In Proceedings of the ACM International Conference on Multimedia, pp. 869-876, New York, USA, October 2004
- [3] N. Shivakumar and H. Garcia-Molina. "Finding near-replicas of documents and servers on the web." In Proceedings of Workshop on Web Databases, pp. 204-212, London, UK, February 1998
- [4] A. Swaminathan, Y. Mao, M. Wu. "Robust and secure image hashing." IEEE Transactions on Information Forensics and Security, Vol. 1, no.2, pp. 215-230, June 2006
- [5] B. Wang, Z. Li, M. Li, and W.-Y. Ma, "Large-scale duplicate detection for Web image search." IEEE International Conference on Multimedia and Expo, pp. 353-356, Toronto, Ontario, Canada, July 2006
- [6] S. Xiang, H.J. Kim, and J. Huang. "Histogram-based image hashing scheme robust against geometric deformations." In Proceedings of the ACM Multimedia and Security Workshop, pp. 121-128, Dallas, Texas, USA, September 2007
- [7] D. Q. Zhang and S.-F. Chang. "Detecting image near-duplicate by stochastic attributed relational graph matching with learning." In Proceedings of the ACM Multimedia, pp. 877-884, New York, USA, October 2004