

■ 2008년도 학생논문 경진대회 수상작

시계열 이동평균 변환을 이용한 노이즈 제어 윤곽선 이미지 매칭

(Noise Control Boundary Image Matching Using Time-Series Moving Average Transform)

김 범 수 [†] 문 양 세 ^{**} 김 진 호 ^{**}
 (Bum-Soo Kim) (Yang-Sae Moon) (Jin-Ho Kim)

요 약 본 논문에서는 윤곽선 이미지 매칭에서 노이즈 제거 정도를 제어하기 위해 시계열 매칭의 이동평균 변환을 이용한다. 이동평균 변환을 윤곽선 이미지 매칭에 적용하게 된 동기는 이동평균 변환이 시계열의 노이즈를 감소시키므로, 이를 사용하면 윤곽선 이미지 매칭에서도 노이즈 제어 효과를 얻을 수 있을 것이라는 직관에 기반한다. 본 논문에서는 우선 윤곽선 이미지 매칭에 이동평균 변환을 적용한 k -계수 이미지 매칭(k -order image matching)을 제안한다. 제안한 k -계수 이미지 매칭은 윤곽선 이미지가 변환된 시계열에 k -이동평균 변환을 적용하여 시계열(이미지) 간의 유사성을 판단한다. 다음으로, 대용량 이미지 데이터베이스를 대상으로 k -계수 이미지 매칭을 수행하기 위한 인덱스 기반 매칭 방법을 제안하고, 그 정확성을 정형적으로 증명한다. 또한, 계수 k 와 매칭 결과와의 관계를 정형적으로 분석하고, 이에 기반하여 계수 k 를 변화시키면서 노이즈 제거 정도를 제어하는 방안을 제시한다. 실험 결과, k -계수 이미지 매칭이 노이즈 제거 효과를 가짐을 확인하였으며, 제안한 인덱스 기반 매칭 방법은 순차 스캔에 비해 수 배에서 수십 배 빠른 성능을 보이는 것으로 나타났다.

키워드 : 시계열 데이터베이스, 데이터 마이닝, 윤곽선 이미지 매칭, 시계열 매칭, 이동평균 변환

Abstract To achieve the noise reduction effect in boundary image matching, we use the *moving average transform* of time-series matching. Our motivation is based on an intuition that using the moving average transform we may exploit the noise reduction effect in boundary image matching as in time-series matching. To confirm this simple intuition, we first propose *k-order image matching*, which applies the moving average transform to boundary image matching. A boundary image can be represented as a sequence in the time-series domain, and our *k-order image matching* identifies similar images in this time-series domain by comparing the *k*-moving average transformed sequences. Next, we propose an index-based matching method that efficiently performs *k-order image matching* on a large volume of image databases, and formally prove the correctness of the index-based method. Moreover, we formally analyze the relationship between an order *k* and its matching result, and present a systematic way of controlling the noise reduction effect by changing the order *k*. Experimental results show that our *k-order image matching* exploits the noise reduction effect, and our index-based matching method outperforms the sequential scan by one or two orders of magnitude.

Key words : Time-series databases, data mining, boundary image matching, time-series matching, moving average transform

· 본 연구는 방위사업청과 국방과학연구소의 지원으로 수행되었습니다.

† 학생회원 : 강원대학교 컴퓨터과학과
bskim@kangwon.ac.kr

** 종신회원 : 강원대학교 컴퓨터학부 컴퓨터과학전공 교수
ysmoon@kangwon.ac.kr
jhkim@kangwon.ac.kr

논문접수 : 2008년 5월 26일
심사완료 : 2008년 7월 30일

Copyright©2009 한국정보과학회 : 개인 목적이거나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.

정보과학회논문지: 데이터베이스 제36권 제4호(2009.8)

1. 서론

컴퓨터의 계산 및 저장 능력이 발전함에 따라, 대용량 시계열 데이터베이스를 대상으로 한 시계열 매칭(time-series matching) 연구가 활발하게 이루어져 왔다[1-4]. 또한, 최근에는 필기체 인식[5], 이미지 매칭[3,6], 바이오 시퀀스 매칭[7] 등의 응용 분야에도 시계열 매칭 연구가 활용되고 있다. 본 논문에서는 이중 대용량 이미지 데이터베이스를 대상으로 하는 윤곽선 이미지 매칭 문제를 다룬다. 윤곽선 이미지 매칭은 이미지의 윤곽선을 시계열로 변환한 후, 시계열 매칭을 사용하여 유사 이미지를 찾는 방법이다[3,6]. 이와 같이 이미지를 시계열로 나타낼 경우, 유사성 판단을 위한 거리 계산과 검색을 위한 인덱스 사용이 용이해져 대용량 데이터베이스 대상의 이미지 매칭이 가능해지게 된다.

본 논문에서는 이동평균 변환을 윤곽선 이미지 매칭에 적용하여, 노이즈에 의한 매칭 결과의 왜곡을 줄이는 방법을 제안한다. 노이즈에 의한 왜곡을 줄이기 위한 전통적인 연구는 이미지 자체에서 노이즈를 제거하는 방법이다[8,9]. 그러나, 이 방법은 노이즈 제거 정도를 제어하기 어렵고, 또한 대용량 데이터베이스에는 적용이 어려운 단점이 있다. 따라서, 본 논문에서는 이미지 자체가 아닌 이미지가 변환된 시계열에서 노이즈를 제거하는 방법을 사용한다.

기존의 시계열 매칭 연구에서는 시계열 자체의 왜곡을 제거하기 위한 방법으로 전처리 변환 기법을 사용하였다[10,11]. 이러한 변환 기법의 예로는 이동평균, 정규화, 쉬프팅 및 스케일링 등이 있다. 본 논문의 연구 동

기는 이러한 변환 기법을 사용하여 시계열의 왜곡을 제거할 경우, 원본 이미지에서도 왜곡이 제거될 것이라는 직관에 기반한다. 즉, 시계열 도메인에서 변환 기법을 적용한다면 이미지 도메인에서도 왜곡 제거의 효과를 거둘 수 있을 것이라는 아이디어에서 출발한다. 특히, 본 논문에서는 시계열 매칭에 있어서 이동평균 변환이 시계열의 노이즈를 제거함에 착안하여, 이를 이미지 매칭에서의 노이즈를 제거하는데 이용하고자 한다.

그림 1은 노이즈를 가진 이미지와 이들의 윤곽선이 이동평균 변환에 의해 어떻게 변경되는지를 나타낸다. 먼저, 그림 1(a)는 노이즈를 가진 이미지들을 나타내며, 그림 1(b)는 이들 윤곽선 이미지가 변환된 시계열을 나타낸다. 다음으로, 그림 1(c)는 그림 1(b)의 시계열들이 8-이동평균 변환된 시계열을 나타낸다. 마지막으로, 그림 1(d)는 그림 1(c)의 이동평균 변환된 시계열을 다시 윤곽선 이미지 형태로 나타낸 것이다. 그림 1을 보면(특히, 그림 1(a)와 1(d)를 비교해 보면), 이동평균 변환이 이미지 윤곽선의 노이즈 제거에 효과가 있음을 확인할 수 있다.

윤곽선 이미지 매칭에서 이동평균 변환을 고려하기 위하여, 본 논문에서는 이동평균 계수 k 의 변화에 따른 k -계수 유사(k -order similar)의 개념을 제안한다. 여기서, 두 윤곽선 이미지가 k -계수 유사하다함은 해당 이미지들이 변환된 두 시계열이 k -이동평균 변환 이후에 유사하다는 의미이다. 그런 다음, k -계수 유사의 개념을 사용하여 윤곽선 이미지 매칭을 수행하는 k -계수 윤곽선 이미지 매칭(간략히 ' k -계수 이미지 매칭'이라 한다)의 개념을 제시하고, 이를 수행하는 효율적인 해결책을 제시한다.

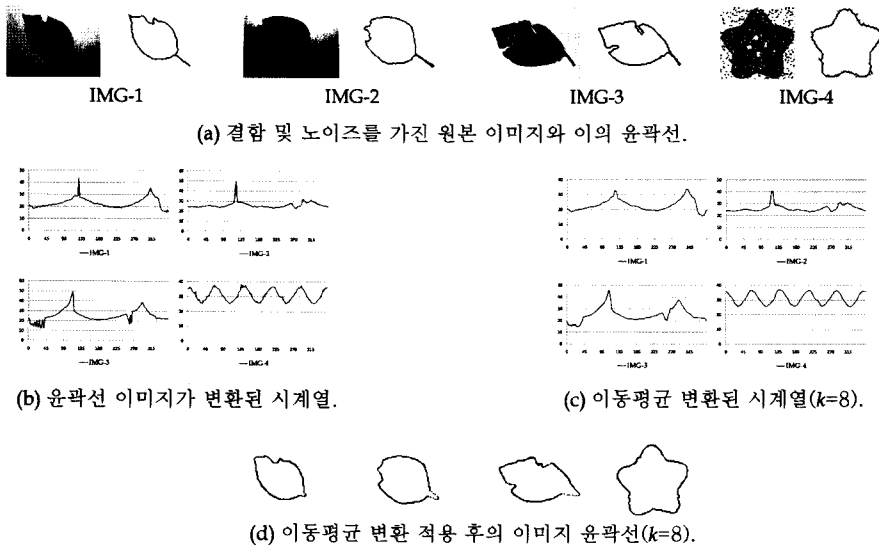


그림 1 이동평균 변환에 의한 이미지 윤곽선의 노이즈 제거 효과

대용량 이미지 데이터베이스에 대한 k -계수 이미지 매칭을 위해 본 논문에서는 인덱스 기반 방법을 제안한다. 이는 기존의 시계열 매칭에서 주로 사용한 방법으로써, 시계열을 저차원 변환한 후 다차원 인덱스를 사용하는 매우 일반적인 방법이다[2,3,10,12]. 본 연구에서는 기존의 저차원 변환[2,4]과 다계수 이동평균 변환(poly-order moving average transform)[10]을 이미지가 변환된 시계열에도 적용할 수 있음을 확인하고, 이를 기반으로 k -계수 이미지 매칭에도 다차원 인덱스를 사용할 수 있음을 정형적으로 증명한다. 또한, 이동평균 계수가 다를 경우 이미지 매칭 결과에 어떤 영향을 미치는지 정형적으로 분석한다. 그 결과, 사용자가 이동평균 계수를 규칙적으로 변화시키면서 노이즈 제거 정도를 제어할 수 있는 방안을 제시한다.

마지막으로 제시한 k -계수 이미지 매칭의 효율성과 제안한 인덱스 기반 매칭 기법의 우수성을 다양한 실험 결과로 확인한다. 이를 위해, k -계수 이미지 매칭을 수행하는 클라이언트-서버 시스템을 구현하여 다양한 종류의 이미지들에 대한 실험을 수행한다. 실험 결과, 계수 k 가 커짐에 따라 질의 이미지와 유사한 많은 이미지들이 정확하게 검색됨을 확인할 수 있었다. 이는 제안한 k -계수 이미지 매칭이 노이즈가 포함된 이미지들에 대한 매칭을 효과적으로 수행함을 의미한다. 결국, 사용자는 계수 k 의 변경을 통해 노이즈 제거 정도를 제어하면서 k -계수 이미지 매칭을 수행할 수 있게 된다. 다음으로, 제안한 인덱스 기반 매칭 기법의 성능이 우수함을 확인하기 위하여 순차 스캔(sequential scan)과의 처리 시간을 비교하였다. 그 결과, 제안한 기법은 순차 스캔에 비해 처리 시간을 수 배에서 수십 배까지 크게 줄인 것으로 나타났다. 이는 제안한 기법이 저차원 변환과 다계수 이동평균 변환을 사용하여 인덱스 기반의 빠른 검색을 수행하기 때문이다. 이 같은 결과를 볼 때, 본 연구는 노이즈 제어 기능을 갖는 윤곽선 이미지 매칭의 실용적 해결책으로 활용될 수 있다고 생각한다.

2. 관련 연구

2.1 시계열 매칭

시계열 데이터는 각 시간별로 측정된 실수 값의 시퀀스이다. 이러한 시계열 데이터의 예로는 주식 데이터, 날씨 데이터, 환율 데이터, 의료 데이터 등이 있다. 시계열 매칭은 사용자에 의해 주어진 질의 시퀀스와 시계열 데이터베이스에 저장된 데이터 시퀀스를 비교하여, 질의 시퀀스와 유사한 데이터 시퀀스를 찾는 작업이다[1-4, 13]. 본 논문에서는 유클리디안(Euclidean) 거리에 기반한 유사 모델[1,2,10]을 사용하여 시계열 매칭을 수행한다. 길이 n 인 두 시퀀스 $X=\{X[1],X[2],\dots,X[n]\}$ 와

$Y=\{Y[1],Y[2],\dots,Y[n]\}$ 의 유클리디안 거리 $D(X,Y)$ 는
$$\sqrt{\sum_{i=1}^n (X[i]-Y[i])^2}$$
 로 정의한다.

두 시퀀스 사이의 거리인 $D(X,Y)$ 가 사용자가 제시한 허용치(tolerance)인 ϵ 이하이면 시퀀스 X 와 Y 는 유사(similar)하다고 정의한다[1,2,4]. 이외에 이동평균 변환[10,11,14], 쉬프팅 및 스케일링[15], 정규화 변환[11] 등의 전처리 변환을 지원하는 유사 모델이 제안되었다. 또한, 유클리디안 거리 대신 DTW(dynamic time warping)[12,13]를 거리 함수로 사용하는 연구가 있었으며, 허용치 ϵ 을 사용한 범위 검색 대신 k -NN(nearest neighbor) 검색[13]에 대한 연구가 진행되었다. 본 논문에서는 이중 가장 많이 연구된 유클리디안 거리와 범위 검색 기반의 시계열 매칭을 윤곽선 이미지 매칭에 활용한다.

본 논문에서는 데이터 시퀀스와 질의 시퀀스의 길이가 동일한 경우의 시계열 매칭인 전체 매칭을 이용한다. 전체 매칭은 Agrawal 등[1]에 의하여 처음 제안되었으며, 크게 전처리, 범위 질의 검색, 후처리의 세 가지 과정으로 구성된다. 전처리 과정은 길이 n 인 데이터 시퀀스를 DFT로 저차원 변환하여, $f(\ll n)$ 개의 특성을 추출한 후, f -차원의 R^* -트리[16]에 저장하는 과정이다. 특성을 추출하는 이유는 다차원 인덱스의 고차원 문제[2, 4]로 인하여, R^* -트리에 고차원인 데이터 시퀀스를 직접 저장하기가 어렵기 때문이다. 저차원 변환을 위한 특성 추출 함수로는 Agrawal 등이 사용한 DFT 이외에도 Wavelet[4], PAA[3,13], SVD[17] 등이 있다. 범위 질의 검색 과정에서는 먼저 질의 시퀀스를 저차원 변환하고, 변환한 점과 허용치 ϵ 으로 범위 질의를 구성한다. 그런 다음, 범위 질의로 R^* -트리를 검색하여 데이터 시퀀스와 유사할 가능성이 높은 후보(candidate)들을 찾아낸다. 저차원 변환의 특성은 착오기각(false dismissal)은 발생하지 않으나, 착오해답(false alarm)이 발생한다[1,2,4]. 따라서, 후보 집합에서 착오해답을 제거하기 위한 후처리 과정을 수행해야 한다. 후처리 과정에서는 질의 시퀀스와 디스크로부터 읽어 온 각 후보 데이터 시퀀스와의 거리를 조사하여 허용치 ϵ 보다 큰 경우는 착오해답으로 제거하고, 그렇지 않은 경우는 최종 유사 시퀀스로 판정한다.

윤곽선 이미지의 노이즈 제거를 위해 본 논문에서는 시계열 이동평균 변환을 사용한다. 이동평균 변환은 시퀀스의 각 엔트리를 평균함으로써 시퀀스 전체의 경향을 파악하는데 매우 유용하다[10,17]. 정확히 말해서, k -이동평균 변환은 주어진 시퀀스에서 연속된 k 개 엔트리의 평균값을 각 엔트리로 하는 새로운 시퀀스를 구성하는 변환이다[14]. 다음 식 (1)은 길이 n 인 시퀀스 $S=\{S[1],\dots,S[n]\}$ 로부터 k -이동평균 변환한 길이 $n-k+1$

인 시퀀스 $S^{(k)} = \{S^{(k)}[1], \dots, S^{(k)}[n-k+1]\}$ 를 구하는 공식이다.

$$S^{(k)}[i] = \frac{1}{k} (S[i] + S[i+1] + \dots + S[i+k-1]) = \frac{1}{k} \sum_{j=i}^{i+k-1} S[j], \quad 1 \leq i \leq n-k+1$$

이동평균 변환에서 k 값은 이동평균 계수(moving average order) 또는 간략히 계수라 하는데, 경향을 파악하고자 하는 주기와 시계열의 노이즈 제거 정도에 따라 달리하여 사용될 수 있다[15].

Rafiei 등[11]은 이동평균 변환을 지원하는 전체 매칭 알고리즘을 제시하였다. 이들은 시퀀스의 길이가 n 일 때, k -이동평균 값을 구할 수 없는 마지막 $(k-1)$ 개의 데이터 값들에 대해서, 부족한 $(k-1)$ 개의 값들을 시퀀스의 맨 앞에서 가져오는 환형적(circular)인 방법을 사용하였다. 다음으로, Moon 등[10]은 하나의 인덱스를 사용하여 임의 계수의 이동평균 변환 서브시퀀스 매칭을 수행하는 방법을 제안하였다. 이들은 시퀀스를 하나의 계수가 아닌 여러 계수로 구성된 계수 집합에 대해서 이동평균 변환하는 다계수 이동평균 변환 개념을 제시하고, 이를 사용하여 임의 계수의 이동평균 변환을 지원하였다. 본 논문에서는 최근의 연구 결과인 Moon 등의 다계수 이동평균 변환을 윤곽선 이미지 매칭에 이용한다.

2.2 이미지 매칭

이미지 매칭[9,18]은 주어진 이미지와 유사한 이미지를 찾는 문제로써 이미지 처리 분야에서의 주요 연구 분야 중 하나이다. 이미지 매칭을 위해 지금까지 다양한 이미지 정보를 이용하려는 시도들이 있었다. 예를 들어, 참고문헌[19]에서는 색상을, 참고문헌[20]에서는 질감을, 참고문헌[21]에서는 모양을 이미지 매칭에 각각 이용하였다. 본 논문에서는 이중 모양 기반의 이미지 매칭에 연구의 초점을 맞춘다. 모양 기반의 이미지 매칭은 윤곽이 있는 객체를 포함한 이미지의 검색하는데 쓰이며, 검

색하고자 하는 이미지의 색상과 질감 등이 비슷한 도메인에 있을 때 효과적이다[9].

모양 기반의 이미지 매칭에서는 주로 객체의 외부 윤곽선이나 영역을 이용한다[21]. 본 논문에서는 이중 외부 윤곽선을 이용하는 CCD(centroid contour distance) 방법[3,6,21]을 활용한다. CCD 방법은 이미지의 외부 윤곽선 중심점에서 n 개 방향의 일정한 각도($\Delta\theta, \Delta\theta = 2\pi/n$)로 나누어 중심점과 외부 윤곽선과의 거리를 계산하여, 이미지를 n -차원 공간의 점으로 매핑하는 방법이다. 그림 2는 CCD 방법에 의해 윤곽선 이미지가 n -차원 공간의 점, 즉 n -차원 시계열로 변환된 예제를 나타낸다. 이와 같이 CCD 방법을 사용하면 윤곽선 이미지를 시계열로 나타낼 수 있고, 이에 따라서 시계열 매칭 기법을 이미지 매칭에 활용할 수 있다[3,6].

이미지의 외부 윤곽선을 시계열로 변환하여 이미지 매칭을 수행하는 기존 연구는 다음과 같다. Vlachos 등[6]은 단순한 이미지 매칭에서 벗어나 인덱스를 사용한 회전 불변 이미지 매칭을 구현하였다. 이 연구에서는 DFT의 진폭이 회전 불변의 특성을 가짐을 보이고, 이를 이미지 매칭에 사용하였다. Keogh 등[3]은 회전 불변 이미지 매칭에 LB_Keogh[12]가 적용됨을 보였다. 하지만, Vlachos 등의 방법은 유사성 검색에 있어서 정보 손실이 있는 거리 척도를 사용하여 매칭의 정확도가 떨어지고 인덱스 구현이 복잡한 단점이 있다. 그리고, Keogh 등의 방법은 인덱스 보다는 메모리 기반 매칭에 적합하다. 또한, 이들 방법은 노이즈가 있는 이미지를 처리하는 방법을 해결하지 않은 점이 본 연구와 차이가 있다.

3. 이동평균 변환 기반 윤곽선 이미지 매칭

3.1 개념

시계열 매칭으로 윤곽선 이미지 매칭을 수행하기 위해서, 우선 윤곽선 이미지 간의 유사성과 이에 기반한 윤곽선 이미지 매칭을 다음과 같이 각각 정의한다.

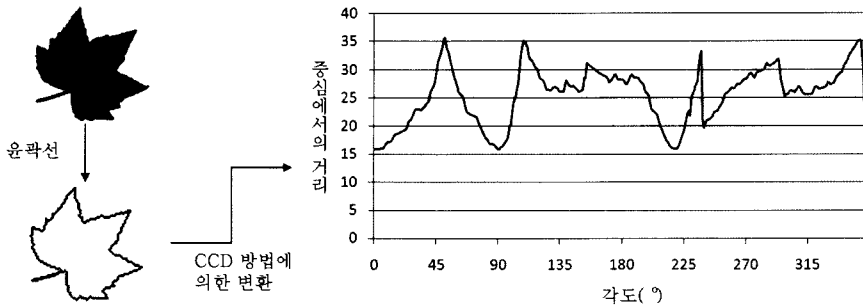


그림 2 CCD 방법을 사용한 이미지의 시계열 변환 예제

정의 1: 두 윤곽선 이미지 A, B 가 CCD 방법으로 변환한 길이 n 인 시계열을 각각 시퀀스 X 와 Y 라 하자. 이때, 두 시퀀스 X, Y 의 유클리디안 거리 $D(X, Y)$ 가 사용자에게 의해 주어진 허용치 ϵ 이하이면 윤곽선 이미지 A 와 B (혹은 시퀀스 X 와 Y)는 유사(similar)하다.

정의 2: 데이터베이스에 저장된 윤곽선 이미지를 CCD 방법으로 변환한 시계열을 ‘데이터 시퀀스(data sequence)’라 하며, 이들을 저장한 데이터베이스를 ‘이미지 시계열 데이터베이스(image time-series databases)’라 한다. 또한, 사용자에게 의해 주어진 질의 윤곽선 이미지를 CCD 방법으로 변환한 시계열을 ‘질의 시퀀스(query sequence)’라 한다. 이때, ‘윤곽선 이미지 매칭(boundary image matching)’이란 질의 시퀀스와 허용치가 주어졌을 때, 질의 시퀀스와 유사한 데이터 시퀀스들을 이미지 시계열 데이터베이스로부터 찾는 작업이다.

그런데, 앞서 제1장에서 설명한 바와 같이 노이즈 제거 효과를 거두기 위해서는 이동평균 변환이 가미된 이미지 매칭을 수행하여야 한다. 즉, 정의 1과 정의 2에서 제시한 윤곽선 이미지 매칭은 이동평균 변환에 의한 노이즈 제거의 개념을 고려하지 않은 문제점이 있다. 특히, 이동평균 계수 k 의 변화에 따라 노이즈 제거 정도가 달라질 수 있으므로, 계수 k 를 고려하여 윤곽선 이미지 간의 유사성을 다시 정의할 필요가 있다. 이를 위해, 우선 기존의 이동평균 변환 정의를 확장하여 윤곽선 이미지가 변환된 시계열의 k -이동평균 변환을 다음과 같이 새롭게 정의한다.

정의 3: 윤곽선 이미지 A 가 변환된 길이 n 인 시퀀스를 $X = \{X[1], \dots, X[n]\}$ 이라 했을 때, X 를 ‘ k -이동평균 변환’한 시퀀스 $X^{(k)} = \{X^{(k)}[1], \dots, X^{(k)}[n]\}$ 은 다음 식 (2)와 같이 정의한다.

$$X^{(k)}[i] = \frac{1}{k}(X[i\%n] + X[(i+1)\%n] + \dots + X[(i+k-1)\%n]) \\ = \frac{1}{k} \sum_{j=i}^{i+k-1} X[j\%n] \quad (2)$$

여기서, $1 \leq i \leq n$, $1 \leq k \leq n-1$ 이고, ‘%’는 나머지 연산자(modular operator)이다.

정의 3의 이동평균 변환은 Rafiei 등[11]이 사용한 환형적 방법과 동일한 정의이다. 이미지의 윤곽선은 끝 부분이 다시 처음 부분으로 연결되며, 이는 시퀀스에 있어서 마지막 엔트리 다음에 다시 처음 엔트리로 연결됨을 의미한다. 따라서, 정의 3과 같이 환형적 방법에 따라 이미지 시퀀스에 대한 이동평균 변환을 정의한다.

다음으로, 이동평균 변환을 고려하여 윤곽선 이미지들의 유사성을 판단하기 위해서, 다음과 같이 윤곽선 이미지 간의 k -계수 유사성을 정의한다.

정의 4: 윤곽선 이미지 A, B 에 대응하는 길이 n 인 시퀀스가 각각 X 와 Y 이고, 이들 X, Y 가 k -이동평균 변환된 시퀀스가 각각 $X^{(k)}, Y^{(k)}$ 라 하자. 이때, 두 시퀀스 $X^{(k)}$ 와 $Y^{(k)}$ 의 유클리디안 거리가 사용자가 제시한 허용치 ϵ 이하이면, 윤곽선 이미지 A 와 B 는 ‘ k -계수 유사(k -order similar)’하다고 정의한다. 즉, 다음 식 (3)이 성립하면, A 와 B (혹은 X 와 Y)는 k -계수 유사하다.

$$D(X^{(k)}, Y^{(k)}) = \sqrt{\sum_{i=1}^n (X^{(k)}[i] - Y^{(k)}[i])^2} \leq \epsilon \quad (3)$$

여기서, $1 \leq k \leq n-1$ 이다.

기존의 윤곽선 이미지 매칭[3,6]에서는 두 시퀀스의 거리 자체를 유사성 판단에 직접 사용하였다. 반면에, 정의 4에서는 이동평균 변환을 고려하기 위하여, 두 시퀀스를 k -이동평균 변환한 이후의 시퀀스간 거리를 유사성 판단의 척도로 사용함을 알 수 있다. 결국, 이동평균 변환을 통해 노이즈를 제거한 이후의 시퀀스들을 유사성 판단에 사용하는 것이다.

정의 4의 k -계수 유사성의 개념에 기반하여, 정의 2의 윤곽선 이미지 매칭을 확장하여 정의한다. 다시 말해서, 정의 2의 윤곽선 이미지 매칭이 이동평균 변환을 고려하지 않은 반면에, 다음의 정의 5는 이를 고려하여 확장한 개념이다.

정의 5: 질의 이미지가 변환된 질의 시퀀스 Q , 허용치 ϵ , 이동평균 계수 k 가 주어졌을 때, 이미지 시계열 데이터베이스로부터 질의 시퀀스 Q 와 k -계수 유사한 데이터 시퀀스들을 찾는 작업을 ‘ k -계수 윤곽선 이미지 매칭(k -order boundary image matching)’이라 정의한다.

혼란이 없는 한 향후 기술에서는 k -계수 이미지 매칭을 k -계수 윤곽선 이미지 매칭과 같은 의미로 사용한다.

3.2 인덱스 구축 및 이미지 매칭 알고리즘

지금까지 설명한 k -계수 이미지 매칭을 사용하면 이동평균 변환을 고려한 매칭이 수행되므로, 결국 노이즈 제거 효과를 볼 수 있다. 그러나, 시계열 매칭의 많은 연구가 그러하듯이, 대용량 이미지 시계열 데이터베이스를 대상으로 한 빠른 검색을 위해서는 인덱스 사용이 필수적 요건이다. 또한 임의의 이동평균 계수를 효율적으로 지원하기 위한 방안이 필요하다. 계수 k 는 노이즈 제거 정도를 제어할 수 있는 변수로써, 이미지의 종류, 사용자의 취향 등에 의해서 변경 가능해야 하기 때문이다. 이와 같이 임의 계수의 k 를 지원하기 위한 가장 단순한 방법은 모든 가능한 k 에 대해서 다차원 인덱스를 구성하는 것이다. 하지만, 이 방법은 여러 인덱스를 위한 많은 저장 공간을 요구할 뿐 아니라, 이미지의 추가, 삭제 시에 여러 인덱스를 갱신해야 하는 어려움이 따른다. 본 논문에서는 기존의 시계열 매칭 연구 결과를 사

용하여 인덱스 사용에 따른 이들 문제를 해결한다.

먼저, 시계열 매칭에서의 저차원 변환은 k -계수 이미지 매칭에서도 그대로 사용할 수 있다. 즉, 저차원 변환 F 가 길이 n 인 시퀀스 X 를 길이 f 인 시퀀스 X_f 로 변환하며, 두 시퀀스 X 와 Y 의 유클리디안 거리에 대해 다음 식 (4)의 하한 조건을 만족하면, 변환 F 는 이미지가 변환된 시퀀스 X, Y 에 대해서도 다음 식 (5)의 하한 조건을 만족한다.

$$D(X, Y) \geq D(X_f, Y_f) = \sqrt{\sum_{i=1}^f (X_f[i] - Y_f[i])^2} \quad (4)$$

$$D(X^{(k)}, Y^{(k)}) \geq D(X_f^{(k)}, Y_f^{(k)}) = \sqrt{\sum_{i=1}^f (X_f^{(k)}[i] - Y_f^{(k)}[i])^2} \quad (5)$$

식 (5)에서 $X^{(k)}$ 는 시퀀스 X 가 k -이동평균 변환된 길이 n 의 시퀀스이며, $X_f^{(k)}$ 는 $X^{(k)}$ 가 F 에 의해 저차원 변환된 길이 f 의 시퀀스이다.

다음으로, 임의 계수 지원을 위하여 Moon 등[10]의 다계수 이동평균 변환 개념을 사용한다. Moon 등은 서브시퀀스 매칭에서 임의 계수의 이동평균 변환을 지원하기 위하여 다계수 이동평균 변환을 사용하였다. 본 논문에서는 이를 서브시퀀스가 아닌 이미지의 시계열에 맞도록 다음과 같이 변경하여 정의한다.

정의 6: 이미지가 변환된 시퀀스 X 를 이동평균 계수의 집합 $\mathbb{K} = \{k_1, k_2, \dots, k_m\}$ 으로 다계수 이동평균 변환한 집합 $X^{(\mathbb{K})}$ 는 $\{X^{(k)} \mid 1 \leq i \leq m\}$ 의 m 개 시퀀스 집합으로 정의된다.

정의 6에 따르면, 다계수 이동평균 변환이란 주어진 시퀀스 X 가 계수 k_i 각각에 의해 k_i -이동평균 변환되어 시퀀스 $X^{(k)}$ 들의 집합을 구성하는 변환임을 알 수 있다.

Moon 등[10]은 다계수 이동평균 변환을 사용할 경우 서브시퀀스 매칭에서 착오기각이 발생하지 않음을 증명하였다. 이에 따라, 이동평균 계수 k 가 집합 \mathbb{K} 의 원소 ($k \in \mathbb{K}$)라 했을 때, 이미지가 변환된 길이 n 의 시퀀스 X, Y 에 대해서도 다음 식 (6)의 하한 조건이 성립한다.

$$D(X^{(\mathbb{K})}, Y^{(\mathbb{K})}) \geq D(X^{(k)}, MBR(Y^{(\mathbb{K})})) \quad (6)$$

식 (6)에서 $MBR(Y^{(\mathbb{K})})$ 는 $Y^{(\mathbb{K})}$ 에 포함된 모든 시퀀스를 포함하는 MBR(minimum bounding rectangle)이다. 결국, 각각의 이동평균 계수 대신에 여러 계수의 집합을 사용함으로써 하나의 인덱스로 임의 계수의 이동평균 변환을 지원할 수 있게 된다.

다음 정리 1은 저차원 변환과 다계수 이동평균 변환을 사용하는 인덱스 기반 매칭 기법이 착오기각을 발생하지 않음을 보이는 이론적 근거이다.

정리 1: 이동평균 계수 k 가 집합 \mathbb{K} 의 원소, 즉

$k \in \mathbb{K}$ 라 하자. 그리고, 저차원 변환 F 가 유사 시퀀스 매칭의 하한 조건을 만족한다고 하자. 그러면, 이미지가 변환된 길이 n 의 시퀀스 X, Y 에 대해서 다음 식 (7)의 하한 조건이 성립한다.

$$D(X^{(k)}, Y^{(k)}) \geq D(X_f^{(k)}, MBR(Y_f^{(\mathbb{K})})) \quad (7)$$

여기서, $Y_f^{(\mathbb{K})}$ 는 $Y^{(\mathbb{K})}$ 에 포함된 길이 n 인 시퀀스들을 F 로 저차원 변환한 길이 f 인 시퀀스들의 집합이며, $MBR(Y_f^{(\mathbb{K})})$ 는 $Y_f^{(\mathbb{K})}$ 의 시퀀스들을 포함하는 f -차원 MBR이다.

증명: 식 (5)에 의해 다음 식 (8)이 성립하고, 식 (6)에 의해 다음 식 (9)가 각각 성립한다.

$$\forall k \in \mathbb{K}, D(X^{(k)}, Y^{(k)}) \geq D(X_f^{(k)}, Y_f^{(k)}) \quad (8)$$

$$D(X_f^{(k)}, Y_f^{(k)}) \geq D(X_f^{(k)}, MBR(Y_f^{(\mathbb{K})})) \quad (9)$$

따라서, 이들 두 식 (8)과 (9)에 의해서 자연히 식 (7)이 성립한다.

정리 1에 따르면, 만일 두 시퀀스 X, Y 가 k -계수 유사하다면, X 가 변환된 f -차원 공간의 점 $X_f^{(k)}$ 와 Y 가 변환된 f -차원 공간의 MBR $MBR(Y_f^{(\mathbb{K})})$ 의 거리가 허용치 ϵ 이하가 된다. 다시 말해서, f -차원 인덱스 공간에서 질의로 주어진 $X_f^{(k)}$ 와 인덱스에 저장된 $MBR(Y_f^{(\mathbb{K})})$ 의 거리가 ϵ 이하일 때, Y 를 X 의 후보 시퀀스로 삼으면 착오기각 없이 k -계수 유사한 이미지를 모두 찾을 수 있다.

그림 3은 인덱스 구축 알고리즘이다. 알고리즘의 입력은 데이터 시퀀스들을 저장한 이미지 시계열 데이터베이스와 지원이 필요한 이동평균 계수들의 집합이다. 알고리즘에서는 각 데이터 시퀀스를 다계수 이동평균 변환 및 저차원 변환하여 다차원 색인에 저장한다. 먼저, 라인 (2)에서는 데이터 시퀀스 Y 를 다계수 이동평균 변환하여 n -차원 시퀀스들의 집합을 구성한다. 다음으로, 라인 (3)에서는 변환된 시퀀스 각각을 f -차원 시퀀스로 저차원 변환한다. 라인 (4)에서는 이들 저차원 변환된 시퀀스들을 포함하는 f -차원 MBR을 구성한다. 그리고, 라인 (5)에서는 구성한 MBR을 시퀀스 Y 의 식별자인 $Y-ID$ 와 함께 다차원 인덱스에 저장한다. 이러한 라인 (2)에서 (5)의 과정을 각각의 데이터 시퀀스에 대해서 반복하여 인덱스를 구성한다(라인 (1)과 (6)). 결국, 그림 3의 알고리즘에 의해 각각의 데이터 시퀀스는 하나의 저차원 MBR로 매핑되어 다차원 인덱스에 저장된다.

그림 4는 앞서 구성한 인덱스를 사용하여 k -계수 이미지 매칭을 수행하는 알고리즘을 나타낸다. 알고리즘의 입력은 질의 이미지가 변환된 질의 시퀀스 X 와 사용자에 의해 주어진 허용치 ϵ 및 이동평균 계수 k 이다. 알

Procedure BuildIndex(Image time-series database DB , A set of orders \mathbb{K})

// Assume each data sequence Y in DB has n dimensions.

- (1) for each data sequence Y in DB do
- (2) Make a set $Y^{(k)}$ of n -dimensional sequences by using the *poly-order moving average transform* on \mathbb{K} ;
- (3) Construct a set $Y_f^{(k)}$ of f -dimensional sequences by using the transformation F ;
- (4) Construct an f -dimensional MBR $MBR(Y_f^{(k)})$ by bounding all f -dimensional sequences;
- (5) Make a record $\langle Y-ID, MBR(Y_f^{(k)}) \rangle$, and store it into the index;
- (6) end for

그림 3 다차원 인덱스 구축 알고리즘

Procedure k-OrderImageMatching(Query sequence X , Tolerance ϵ , Order k)

// Assume the query sequence X has n dimensions.

- (1) Make an n -dimensional sequence $X^{(k)}$ from X by using the *k-order moving average transform*;
- (2) Transform $X^{(k)}$ to an f -dimensional sequence $X_f^{(k)}$ by using the transformation F ;
- (3) Make a range query using $X_f^{(k)}$ and ϵ ;
- (4) Construct a candidate set by evaluating the range query on the index;
- (5) Identify k -order similar images from the candidate set through the post-processing step;

그림 4 k-계수 이미지 매칭 알고리즘

고리즘에서는 다차원 인덱스를 검색하여 후보 시퀀스들을 구하고, 실제 데이터베이스를 액세스하여 k -계수 유사한 시퀀스만을 식별해 낸다. 먼저, 라인 (1)에서는 n -차원의 질의 시퀀스 X 를 k -이동평균 변환하여 n -차원 시퀀스인 $X^{(k)}$ 를 구성한다. 다음으로, 라인 (2)에서는 이를 F 로 저차원 변환하여 f -차원 시퀀스 $X_f^{(k)}$ 를 만든다.

라인 (3)에서는 f -차원 시퀀스 $X_f^{(k)}$ 와 허용치 ϵ 을 사용하여 f -차원 공간의 범위 질의를 구성한다. 그리고, 라인 (4)에서는 범위 질의로 앞서 구축한 다차원 인덱스를 검색하여 k -계수 유사할 가능성이 높은 후보 시퀀스들을 찾아낸다. 이렇게 구성한 후보 집합에는 k -계수 유사 이미지와 함께 착오해답이 포함되어 있을 수 있다. 따라서, 마지막 단계인 라인 (5)에서는 실제 데이터 시퀀스를 액세스하고 거리 계산을 수행하여 k -계수 유사 이미지만을 식별해내는 후처리 과정을 수행한다.

3.3 이동평균 계수와 매칭 결과와의 관계

본 절에서는 이동평균 계수와 매칭 결과 사이의 관계를 논의한다. 먼저, 두 이동평균 계수 k_1 과 k_2 에서 $k_2 = 2k_1$ 관계가 성립하면, 두 계수의 매칭 결과 사이에는 다음의 포함 관계가 성립한다.

정리 2: 이동평균 계수 k_1 과 k_2 가 주어졌고, $k_2 = 2k_1$ 의 관계가 성립한다 하자. 그러면, k_2 -계수 이미지 매칭의 결과는 항상 k_1 -계수 이미지 매칭 결과를 포함한다.

즉, k_1 -계수 이미지 매칭 결과가 k_2 -계수 이미지 매칭 결과의 부분 집합이다.

증명: k_1 -계수 이미지 매칭 결과가 k_2 -계수 이미지 매칭 결과에 포함됨을 보이기 위해서는 두 시퀀스 X 와 Y 에 대해 다음 식 (10)이 성립함을 보이면 된다.

$$D(X^{(k_1)}, Y^{(k_1)}) \geq D(X^{(k_2)}, Y^{(k_2)}) \quad (10)$$

식 (10)의 의미는 X 와 Y 가 k_1 -계수($D(X^{(k_1)}, Y^{(k_1)}) \leq \epsilon$)

유사하다면 이들 둘은 반드시 k_2 -계수 유사($D(X^{(k_2)}, Y^{(k_2)}) \leq \epsilon$)하다는 의미이다. 식 (10)은 다음 과정에 의해 성립한다. 먼저, $X^{(k_1)}$ 와 $Y^{(k_1)}$ 을 다음과 같은 시퀀스로 나타내기로 하자.

$$X^{(k_1)} = \{X^{(k_1)}[1], X^{(k_1)}[2], \dots, X^{(k_1)}[n]\},$$

$$Y^{(k_1)} = \{Y^{(k_1)}[1], Y^{(k_1)}[2], \dots, Y^{(k_1)}[n]\}$$

그러면, 두 시퀀스 $X^{(k_1)}$ 와 $Y^{(k_1)}$ 의 거리 제곱인 은 다음과 같이 구할 수 있다.

$$D^2(X^{(k_1)}, Y^{(k_1)}) = \sum_{i=1}^n (X^{(k_1)}[i] - Y^{(k_1)}[i])^2$$

다음으로, $X^{(k_2)}$ 와 $Y^{(k_2)}$ 를 다음과 같은 시퀀스로 나타내기로 하자.

$$X^{(k_2)} = \{X^{(k_2)}[1], X^{(k_2)}[2], \dots, X^{(k_2)}[n]\},$$

$$Y^{(k_2)} = \{Y^{(k_2)}[1], Y^{(k_2)}[2], \dots, Y^{(k_2)}[n]\}$$

조건에 의해, $k_2 = 2k_1$ 의 관계가 성립하므로, 시퀀스 $X^{(k_2)}$ 와 $Y^{(k_2)}$ 는 다음과 같이 $X^{(k_1)}$ 과 $Y^{(k_1)}$ 의 원소를 사용하여 나타낼 수 있다.

$$X^{(k_2)} = \left\{ \frac{X^{(k_1)}[1] + X^{(k_1)}[1+k_1]}{2}, \frac{X^{(k_1)}[2] + X^{(k_1)}[2+k_1]}{2}, \dots, \frac{X^{(k_1)}[n] + X^{(k_1)}[(n+k_1)\%n]}{2} \right\},$$

$$Y^{(k_2)} = \left\{ \frac{Y^{(k_1)}[1] + Y^{(k_1)}[1+k_1]}{2}, \frac{Y^{(k_1)}[2] + Y^{(k_1)}[2+k_1]}{2}, \dots, \frac{Y^{(k_1)}[n] + Y^{(k_1)}[(n+k_1)\%n]}{2} \right\}$$

그러면, 두 시퀀스 $X^{(k_2)}$ 와 $Y^{(k_2)}$ 의 거리 제곱인 $D^2(X^{(k_2)}, Y^{(k_2)})$ 는 다음과 같이 구할 수 있다.

$$D^2(X^{(k_2)}, Y^{(k_2)}) = \frac{1}{4} \left(\sum_{i=1}^n 2(X^{(k_1)}[i] - Y^{(k_1)}[i])^2 + \sum_{i=1}^n 2(X^{(k_1)}[i] - Y^{(k_1)}[i]) (X^{(k_1)}[(i+k_1)\%n] - Y^{(k_1)}[(i+k_1)\%n]) \right)$$

앞서 구한 두 거리의 제곱에 대해 다음의 관계식이 성립한다.

$$4(D^2(X^{(k_1)}, Y^{(k_1)}) - D^2(X^{(k_2)}, Y^{(k_2)})) = \sum_{i=1}^n 2(X^{(k_1)}[i] - Y^{(k_1)}[i])^2 - \sum_{i=1}^n 2(X^{(k_1)}[i] - Y^{(k_1)}[i]) (X^{(k_1)}[(i+k_1)\%n] - Y^{(k_1)}[(i+k_1)\%n])$$

여기서,

$$\sum_{i=1}^n (X^{(k_1)}[i] - Y^{(k_1)}[i])^2 = \sum_{i=1}^n (X^{(k_1)}[(i+k_1)\%n] - Y^{(k_1)}[(i+k_1)\%n])^2$$

이 성립하므로, 상기 식은 다음과 같이 바꾸어 쓸 수 있다.

$$4(D^2(X^{(k_1)}, Y^{(k_1)}) - D^2(X^{(k_2)}, Y^{(k_2)})) = \sum_{i=1}^n ((X^{(k_1)}[i] - Y^{(k_1)}[i]) - (X^{(k_1)}[(i+k_1)\%n] - Y^{(k_1)}[(i+k_1)\%n]))^2$$

상기 식에서 최종 우변은 항상 0 이상이므로 두 거리의 제곱 사이에는 다음 관계식이 성립한다.

$$D^2(X^{(k_1)}, Y^{(k_1)}) - D^2(X^{(k_2)}, Y^{(k_2)}) \geq 0 \Leftrightarrow$$

$$D^2(X^{(k_1)}, Y^{(k_1)}) \geq D^2(X^{(k_2)}, Y^{(k_2)})$$

그러면, 시퀀스 간의 거리 $D(X^{(k_1)}, Y^{(k_1)})$ 및 $D(X^{(k_2)}, Y^{(k_2)})$

는 항상 0 이상이므로, 상기 조건에 의해 결국 식 (10)이 성립한다.

정리 2의 의미는 k -계수 이미지 매칭을 수행할 때, 하나의 계수에 대한 결과는 이를 두 배한 계수의 결과에 항상 포함됨을 의미한다. 이는 또한 노이즈를 고려하지 않은 이미지 매칭 결과($k=1$ 인 경우의 결과)가 노이즈를 고려한 매칭 결과($k=2, 4, 8, \dots$ 인 경우의 결과)에 항상 포함됨을 보장한다. 즉, 제안한 k -계수 이미지 매칭은 기존의 단순한 이미지 매칭을 포함하는 보다 일반적인 매칭 기법이라 할 수 있다. 결국, 사용자는 작은 계수(예를 들어, 1이나 2)로부터 매칭을 시작하여 계수를 두 배씩 크게 함으로써, 노이즈 제거 정도를 제어하면서 대화식(interactive) 이미지 매칭을 수행할 수 있게 된다.

상기 정리 2와 같이 $k_2 = 2k_1$ 의 두 배 관계가 아닌 경우에 대해서는 일반적인 포함 관계를 밝히지 못하였다. 특히, $k_1 < k_2$ 인 일반적인 관계에서, k_1 -계수 이미지 매칭 결과가 k_2 -계수 이미지 매칭 결과에 포함되지는 않는 것으로 나타났다. 다음 예제 1이 이러한 반례를 나타낸다.

예제 1: 두 시퀀스가 $X = \{2, 4, 2, 4\}$, $Y = \{1, 1, 1, 1\}$ 로 주어졌다 하자. 그러면, 2-이동평균 변환된 시퀀스는 각각 $X^{(2)} = \{3, 3, 3, 3\}$, $Y^{(2)} = \{1, 1, 1, 1\}$ 로 계산되며, 3-이동평균 변환된 시퀀스는 각각 $X^{(3)} = \{8/3, 10/3, 8/3, 10/3\}$, $Y^{(3)} = \{1, 1, 1, 1\}$ 로 계산된다. 여기서, 두 시퀀스 간의 거리를 각각 구하면 다음 식 (11)과 같다.

$$D(X^{(1)}, Y^{(1)}) = \sqrt{20}, \quad D(X^{(2)}, Y^{(2)}) = \sqrt{16},$$

$$D(X^{(3)}, Y^{(3)}) = \sqrt{\frac{148}{3}} \approx \sqrt{16.44} \quad (11)$$

식 (11)의 결과를 보면, $D(X^{(1)}, Y^{(1)})$ 과 $D(X^{(2)}, Y^{(2)})$ 사이에는 정리 2의 관계와 같이 $\sqrt{20} \geq \sqrt{16}$ 이 성립함을 알 수 있다. 반면에, $D(X^{(2)}, Y^{(2)})$ 와 $D(X^{(3)}, Y^{(3)})$ 사이에는 $\sqrt{16} \geq \sqrt{\frac{148}{3}} (\approx \sqrt{16.44})$ 의 관계가 성립하지 않는다.

즉, 계수 2와 3 사이에는 $2 \leq 3$ 의 관계가 성립하지만, 2-계수 이미지 매칭 결과가 3-계수 이미지 매칭 결과에 포함되지 않을 수 있음을 의미한다.

예제 1은 서로 다른 두 계수 k_1 과 k_2 를 모두 아우르는 일반적인 포함 관계를 찾기는 어려움을 의미한다. 따라서, 본 논문에서는 두 계수 k_1 과 k_2 가 두 배 관계인 경우의 포함 관계만을 정리 2에서 정형적으로 증명하였다.

그런데, 대부분의 경우에 있어서 $k_1 < k_2$ 이면 k_2 -계수 이미지 매칭 결과가 k_1 -계수 이미지 매칭 결과에 포함 되는 것으로 나타났다. 정리 2의 특수한 경우에는 이 관계가 명확하며, 두 배 조건이 아닌 경우의 실험 결과에서도 대부분 유사한 포함 관계를 얻을 수 있었다. 이는 계수 k 가 증가할수록 노이즈 제거 효과가 커져서 두 이미지가 보다 유사해지는 경향을 보이기 때문이다. 결국, 사용자는 이동평균 계수 값을 증가시키면서(특히, 두 배씩 증가시키면서), 자신이 원하는 노이즈 제어 정도를 제어할 수 있게 된다. 그리고, 이미지나 이 이미지가 변환된 시계열의 종류와 특성에 따라 최적의 계수 k 를 찾는 연구를 고려할 수 있으며, 이러한 최적 계수의 선택은 향후 연구로 남겨둔다.

4. 성능 평가

4.1 실험 데이터 및 환경

본 논문에서는 클라이언트-서버 기반의 k -계수 이미지 매칭 시스템을 구현하여 실험에 사용하였다. 이미지의 시계열 변환 과정은 그레이 스케일(gray scale) 이미지에 대한 이진화 변환과 윤곽선 추적으로 나뉜다. 이와 같은 변환 과정은 이진화 임계값의 정도와 윤곽선 추적 기법에 따라 추출 결과가 다르게 나타나는 단점도 있다. 따라서, 본 논문에서는 이진화 임계값을 240으로 고정하여 사용하였고, 윤곽선 추적 기법은 일반적으로 많이 쓰이는 8-이웃 연결성(8-Neighborhoods Connectivity) 추적 알고리즘을 사용하였다. 또한 본 논문에서는 노이즈 추가 효과를 구현하였다. 구현한 노이즈 효과는 블러(blur) 효과, 랜덤 노이즈 효과, 그리고 이들을 혼합하는 효과가 있다. 먼저, 그림 5(a)는 블러 효과에 대한 화면을 나타낸다. 그림 5(a)의 왼편은 원본 이미지를, 오른편은 블러 효과가 적용된 이미지를 각각 나타낸다. 다음으로, 그림 5(b)는 랜덤 노이즈 효과에 대한 화면이다. 마찬가지로 왼편은 원본 이미지를, 오른편은 랜덤 노

이즈가 추가된 이미지를 나타낸다. 또한, 이들 두 가지 효과를 혼합하여 사용하는 것도 가능하다. 즉, 블러 효과 후에 랜덤 노이즈 효과를 적용하거나, 랜덤 노이즈 효과 후에 블러 효과의 적용이 가능하다. 본 논문에서는 원본 이미지에 추가된 블러 효과, 랜덤 노이즈 효과를 노이즈의 일종으로 생각하고, 이들에 강건한 이미지 매칭을 수행하는 실험으로 수행하였다.

실험에서는 총 9만 개의 이미지를 갖는 이미지 데이터베이스를 구축하였다. 이를 위해 총 1만 개의 원본 이미지를 웹 상에서 직접 구하였다. 그림 6은 이러한 원본 이미지의 예를 나타낸다. 그림에서 보듯이, 하나의 이미지에 하나의 객체만 포함된 경우도 있으나, 하나의 이미지에 두 개 이상의 객체가 포함된 경우도 더러 있는 것으로 나타났다. 다음으로, 실험을 위해 각각의 원본 이미지로부터 여덟 가지의 노이즈 이미지를 인위적으로 생성하였다. 그림 7은 원본 이미지로부터 여덟 가지의 노이즈 이미지를 생성한 과정을 나타낸다. 그림에서 보듯이, 블러 효과, 랜덤 노이즈 효과를 하나만 혹은 둘 모두 적용하여 하나의 원본 이미지에서 총 여덟 개의 노이즈 이미지를 생성하였다. 결국, 원본 이미지 1만 개와 노이즈 이미지 8만 개 등 총 9만 개를 실험에 사용하였다.

원본 및 노이즈 이미지로부터 생성된 시계열 시퀀스는 약 10만 개로, 이미지 시계열 데이터베이스에는 이들 10만 개의 데이터 시퀀스를 저장한다. 앞서 설명한 바와 같이 하나의 이미지가 두 개 이상 객체를 가질 수 있고, 이 경우는 하나의 이미지에서 두 개 이상의 데이터 시퀀스가 추출될 수 있다. 그림 6(b)와 6(d)가 그러한 예이다. 결국, 총 9만 개의 이미지를 사용하였으나, 여기에서 생성된 데이터 시퀀스는 9만 개보다 많은 총 104,340 개가 구축되었다.

성능 실험은 서버에서 수행하였으며, 그 환경은 다음과 같다. 먼저 하드웨어 플랫폼은 UltraSPARC III CPU



그림 5 원본 이미지에 노이즈를 추가한 화면 예제

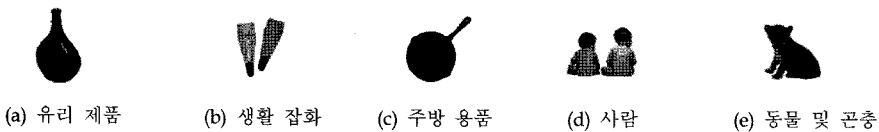


그림 6 원본 이미지 예제

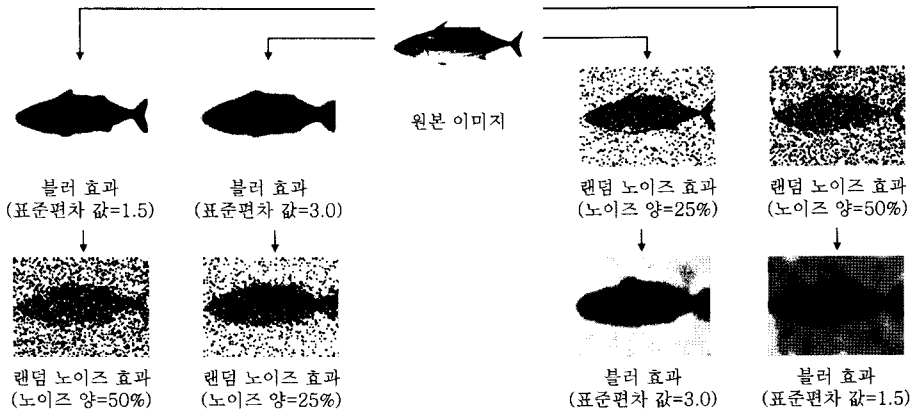


그림 7 원본 이미지에서 생성된 여덟 가지 종류의 노이즈 이미지 예제

1.34GHz, 1.0GB RAM, 80GB 하드 디스크를 장착한 SUN Ultra 25이며, 소프트웨어 플랫폼은 Solaris 10 운영 체제이다. 다차원 인덱스로는 R*-트리를 사용하였으며, 페이지 및 인덱스 페이지의 크기는 4,096 바이트를 사용하였다. 그리고, 각 윤곽선 이미지는 길이 360의 시퀀스로 변환하였으며, 각 시퀀스에서는 고차원 문제를 고려하여 열 개보다 작은 DFT의 여덟 개의 특성을 추출하였다.

4.2 실험 결과

4.2.1 k-계수 이미지 매칭 검색 결과

계수 k의 변화에 따른 k-계수 이미지 매칭의 검색 결과를 확인하기 위하여 다양한 이미지를 사용하여 반복적인 실험을 수행하였다. 먼저, 그림 8은 향아리 모양의 이미지를 질의로 하여 실험한 결과이다. 실험에서 허용치 ϵ 은 이동평균 변환을 하지 않고 이미지 매칭을 수행했을 때(즉, k가 1일 때), 두 개의 이미지가 결과로 반

환되는 값 중 가장 큰 값으로 설정하였다. 예를 들어, 그림 8에서는 ϵ 이 13.5일 때 1-계수 이미지 매칭 결과가 두 개였으며, ϵ 이 13.5보다 크면 결과가 세 개 이상이 되었다. 따라서, $\epsilon=13.5$ 으로 고정하고, k를 변화시키면서 k-계수 이미지 매칭을 수행한 결과이다. 그림을 보면, 계수가 커짐에 따라 질의 이미지와 유사한 이미지들이 많이 검색됨을 알 수 있다. 특히, 향아리의 원본 이미지에 블러 효과, 랜덤 노이즈 효과 등이 추가된 노이즈 이미지들이 k가 증가함에 따라서 유사 이미지로 검색됨을 확인할 수 있다. 이는 제한한 k-계수 이미지 매칭이 노이즈가 포함된 이미지 데이터베이스에 대한 매칭을 효과적으로 수행함을 의미한다.

그림 9는 질의 이미지로 노이즈가 포함된 나뭇잎 이미지를 사용한 매칭 결과이다. 이 실험에서는 원본 나뭇잎 이미지에 랜덤 노이즈 효과(노이즈 양=50%)를 주어

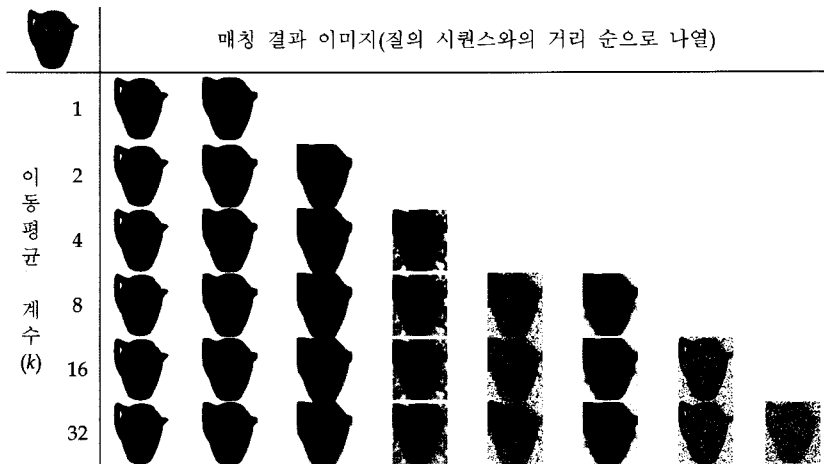


그림 8 향아리 이미지의 k-계수 이미지 매칭 결과(허용치 = 13.5)

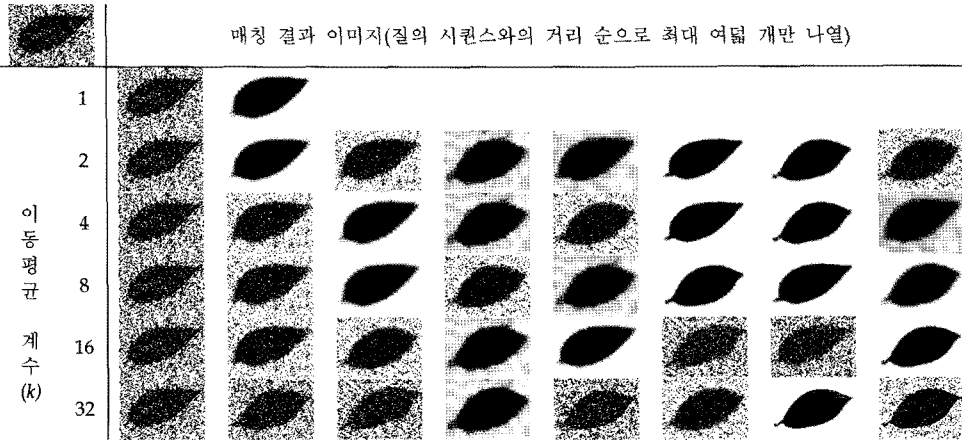


그림 9 랜덤 노이즈 효과(50%)가 적용된 나뭇잎 이미지의 k -계수 이미지 매칭 결과(허용치 = 55.0)

노이즈 이미지를 만들고, 이를 질의 이미지로 사용하였다. 허용치 ϵ 은 그림 8의 실험과 같은 방식에 따라 55.0로 결정하여 사용하였다. 그림에서 보듯이, 이동평균 계수 값이 증가함에 따라 보다 많은 유사 이미지가 검색됨을 알 수 있다. 그러나 그림 8과는 달리, 그림 9의 결과를 보면, k 가 증가함에 따라 유사하지만 원본과 다른 나뭇잎 이미지들도 검색됨을 알 수 있다. (그림에서, k 가 2 혹은 4일 때의 일곱 번째 이미지는 실제로는 질의 이미지와 원본이 다른 것들임에 유의한다.) 이는 k 값의 증가가 노이즈 제거 효과를 가져옴과 동시에 다른 이미지를 유사하다고 판단할 수 있을 가능성도 높임을 의미한다.

표 1은 그림 8 및 그림 9의 실험을 다양한 종류의 다

른 이미지에 수행한 결과를 요약한 것이다. 표 1에서 검색 결과는 k -계수 이미지 매칭을 수행한 결과 나온 이미지들의 개수이고, 유사 이미지는 검색 결과 이미지들 중에서 주어진 질의 이미지와 실제로 유사한 이미지들, 즉 노이즈의 유무만 다르고 본디 같은 원본에서 출발한 이미지들의 개수를 의미한다. 따라서, 유사 이미지 개수가 클수록 노이즈에 무관하게 매칭이 이뤄짐을 의미하며, 검색 결과 대비 유사 이미지의 비율이 클수록 보다 정확한 검색이 이뤄짐을 의미한다. 표를 보면, 모든 이미지에 있어서 k 가 증가함에 따라 검색 결과가 늘어나고, 이들 검색 결과 모두가 유사 이미지로 판별되었음을 알 수 있다. 반면에 앞서 그림 9와 같이 결과 중 일부가 질의 이미지의 원본과 다른 것은 매우 이례적인 예로

표 1 여러 이미지에 대한 k -계수 이미지 매칭 결과 요약

질의 이미지	이름	노이즈 효과	이미지 개수	이동평균 계수 (k)					
				1	2	4	8	16	32
	모자	적용 안함	검색 결과	2	3	4	5	6	7
			유사 이미지	2	3	4	5	6	7
	토끼	적용 안함	검색 결과	2	5	7	7	8	8
			유사 이미지	2	5	7	7	8	8
	사람	블러 효과 (표준편차 값=1.5)	검색 결과	2	5	5	5	6	7
			유사 이미지	2	5	5	5	6	7
	전구	블러 효과 (표준편차 값=3.0)	검색 결과	2	4	4	4	5	7
			유사 이미지	2	4	4	4	5	7
	표지판	랜덤 노이즈 효과 (노이즈 양=25%)	검색 결과	2	3	3	4	5	7
			유사 이미지	2	3	3	4	5	7
	나비	랜덤 노이즈 효과 (노이즈 양=50%)	검색 결과	2	2	3	4	5	6
			유사 이미지	2	2	3	4	5	6

밝혀졌다. 결국 제안한 k -계수 이미지 매칭은 비교적 정확하게 유사 이미지들을 검색한다고 말할 수 있으며, 계수 k 는 노이즈 제거 효과를 제어하는 척도로 사용할 수 있다고 말할 수 있다.

지금까지의 k -계수 이미지 매칭 실험 결과를 요약하면 다음과 같다. 제안한 k -계수 이미지 매칭에서는 계수 k 가 증가할수록 보다 많은 유사 이미지들을 검색해 낸다. 이는 k -계수 이미지 매칭이 노이즈를 제거한 상태에서 매칭을 수행하고, 노이즈 제거 정도는 k 값에 비례하기 때문이다. 그런데 k 값을 너무 크게 주면, 유사하지 않은 이미지들도 매칭 결과에 일부 포함될 수 있다. 이는 실제 유사 이미지뿐만 아니라 그렇지 않은 이미지들도 이동평균 변환에 의한 완만해지(smoothing) 영향을 받기 때문이다. 결국, 사용자는 계수 k 를 조절하면서 자신이 원하는 유사 이미지를 찾는 대화식 접근법이 필요하다.

4.2.2 처리 시간

본 절에서는 k -계수 이미지 매칭을 수행하는데 있어서, 제안한 인덱스 기반 방법의 처리 시간을 순차 스캔과 비교한다. 순차 스캔은 이미지 시계열 데이터베이스의 데이터 시퀀스를 순차적으로 액세스하며 유사 시퀀스를 찾는 단순한 방법이다. 제안한 매칭 방법은 저차원 변환 및 다계수 이동평균 변환을 이용하므로, 순차 스캔에 비해서 우수한 성능을 보이게 된다. 본 절에서는 이러한 성능 개선 효과를 다양한 질의 이미지 집합에 대한 실험으로 확인한다.

실험 결과를 설명하기 위해서 먼저 선택율(selectivity)[2,4,13]을 다음 식 (12)와 같이 정의한다. 식 (12)에서 보듯이, 선택율은 전체 데이터 시퀀스 중에서 몇 개의 시퀀스가 1-계수 이미지 매칭 결과로 반환되는지를 나타낸다. 실험에서는 선택율 10E-4, 10E-3, 10E-2 각각에 대한 허용치들을 결정한 후, 해당 허용치에 대해 k 를 변경하면서 처리 시간을 측정하였다.

$$\text{선택율} = \frac{1\text{-계수 이미지 매칭 결과 검색된 시퀀스 개수}}{\text{시계열 이미지 데이터베이스에 저장된 시퀀스 개수}} \quad (12)$$

질의 이미지로는 원본 집합, 블러 집합, 랜덤 노이즈 집합, 혼합 집합의 네 가지 종류를 사용하였다. 원본 집합은 원본 이미지 20개로, 블러 집합은 블러 효과(1.5, 3.0)를 적용한 노이즈 이미지 20개로, 랜덤 노이즈 집합은 랜덤 노이즈 효과(25%, 50%)를 적용한 노이즈 이미지 20개로 각각 구성하였다. 그리고, 혼합 집합은 이미지 시계열 데이터베이스 전체에서 임의로 선택한 20개의 이미지로 구성된 이미지 집합이다. 실험 결과로는 20개 이미지의 처리 시간 평균 값을 사용하였다. 약 10만개의 시계열에 대한 인덱스 구축에 걸린 시간은 약 $3 \times 10E+8(\mu\text{sec})$ 이며, 이는 임의의 이동평균 계수 k 를 지원하기 위해 모든 가능한 k 에 대한 인덱스를 구축하는데 걸리는 시간과 비교한다면 전처리 과정으로써 무시할만한 오버헤드임을 알 수 있다.

그림 10은 원본 집합에 대한 순차 스캔과 제안한 매칭 방법의 처리 시간을 나타낸다. 그림 10(a)는 선택율이 10E-4인 경우의 실험 결과이고, 그림 10(b)와 10(c)는 각각 10E-3 및 10E-2인 경우의 실험 결과이다. 그림을 보면, 제안한 매칭 방법의 처리 시간이 순차 스캔에 비해 훨씬 작음을 알 수 있다. (그래프의 Y축이 로그 스케일임에 유의한다.) 이러한 성능 개선은 제안한 매칭 방법이 다차원 인덱스를 사용하기 때문으로, 매우 당연한 결과라 할 수 있다. 그림 10을 보면, 계수 k 가 증가할수록 두 방법 모두 처리 시간이 조금씩 늘어남을 알 수 있다. 이는 k 값이 증가할수록 후보 개수가 늘어나고 또한 이동평균 변환 계산에 시간이 오래 걸리기 때문이다. 다음으로, 선택율이 증가할수록 두 방법의 처리 시간 차이가 줄어들음을 알 수 있다. (그림 10(a)에 비해 그림 10(c)에서 두 방법의 처리 시간 차이가 줄어들어 유의한다.) 이는 순차 스캔은 선택율에 무관한 반면에, 제안한 매칭 기법은 선택율이 증가하면 액세스하는 데이터 시퀀스가 많아지기 때문이다. 즉, 순차 스캔의 경우는 선택율에 관계없이 전체 데이터베이스를 액세스해야 하므로 처리 시간이 일정한 반면, 제안한 인덱스 기반 방법은 선택율이 커지면 범위 질의의 검색 범위가

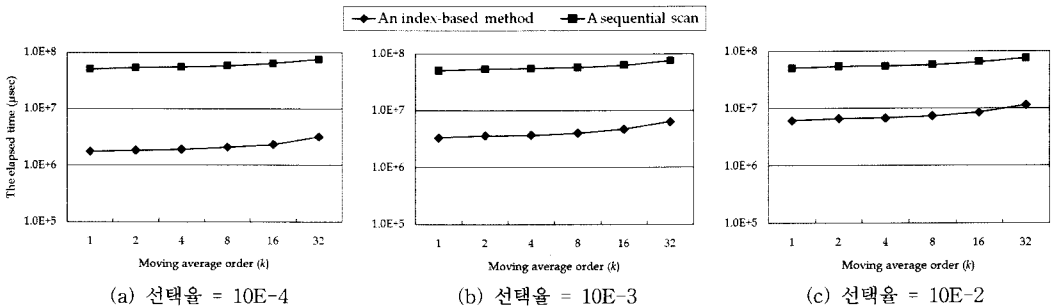


그림 10 원본 집합에 대한 처리 시간 실험 결과

넓어져 많은 후보 시퀀스가 검색되고 이들 중 착오해답의 식별을 위해 많은 데이터 시퀀스를 액세스해야 하기 때문이다. 그런데 이미지 매칭에서는 선택율이 낮은 경우, 즉 수 개에서 수십 개의 이미지만 식별하는 경우가 보다 일반적인 것이다. 따라서, 제안한 인덱스 기반 매칭 방법이 순차 스캔에 비해 훨씬 실용적인 방법이라 할 수 있다. 그림 10의 실험 결과를 요약하면, 제안한 매칭 방법이 순차 스캔에 비해 6.7배에서 29.4배까지 성능을 향상시킨 것으로 나타났다.

다음으로, 블러 집합, 랜덤 노이즈 집합, 혼합 집합에 대한 실험을 수행하였다. 그런데, 그 결과는 그림 10의 원본 집합의 결과와 매우 유사하게 나타났다. (결과가 거의 동일하므로 자세한 실험 결과는 생략한다.) 즉, 다른 이미지 집합을 사용하더라도 제안한 방법이 순차 스캔에 비해 우수한 결과를 나타냈으며, 이는 제안한 방법이 이미지의 종류에 관계없이 우수한 성능을 보임을 의미한다. 결국, 제안한 인덱스 기반 방법은 수십만 개 이상의 이미지로 구성된 대용량 이미지 데이터베이스에 대한 k -계수 이미지 매칭을 수행할 수 있는 효율적인 방법이라 할 수 있다.

5. 결론

본 논문에서는 시계열 매칭의 이동평균 변환을 윤곽선 이미지 매칭에 적용하여 노이즈 제거 효과를 얻는 새로운 이미지 매칭 접근법을 시도하였다. 제안한 접근법은 이동평균 변환이 시계열의 노이즈를 감소시키므로, 이를 이용하면 윤곽선 이미지 매칭에서 노이즈 제거 효과를 얻을 수 있을 것이라는 직관에 기반한다. 제안한 윤곽선 이미지 매칭은 이와 같이 이동평균 변환을 사용함으로써 사용자가 노이즈 제거 정도를 제어하며 매칭을 수행할 수 있으며, 저차원 변환 및 다계수 이동평균에 기반한 다차원 인덱스를 사용함으로써 매우 우수한 매칭 성능을 제공한다.

본 논문의 공헌은 다음과 같이 요약할 수 있다. 첫째, 이동평균 변환을 고려한 k -계수 유사와 k -계수 이미지 매칭의 개념을 새롭게 제안하였다. 기존의 윤곽선 이미지 매칭에서는 이동평균 변환을 고려하지 않은 관계로, 본 논문에서는 기존의 유사성 척도 및 윤곽선 이미지 매칭을 확장하여 k -계수 유사와 k -계수 이미지 매칭을 정형적으로 정의하였다. 둘째, 제안한 k -계수 이미지 매칭에서 다차원 인덱스를 사용한 검색이 가능함을 정리 1에서 정형적으로 증명하였다. k -계수 이미지 매칭은 이동평균 변환의 정의 등에 있어서 기존 시계열 매칭과 차이점이 있으므로, 본 논문에서는 k -계수 이미지 매칭에 기존의 저차원 변환 및 다계수 이동평균 변환을 사용할 수 있음을 차례로 보이고, 이를 통해 다차원 인덱

스를 사용한 검색이 가능함을 증명하였다. 또한, 이러한 이론적 근거를 바탕으로 k -계수 이미지 매칭을 위한 인덱스 구축 및 이미지 매칭 알고리즘을 제안하였다. 셋째, 이동평균 계수 변화에 따른 매칭 결과의 관계를 정리 2에서 정형적으로 설명하였다. 이 관계에 따라 사용자가 이동평균 계수를 변경함으로써 노이즈 제거 정도를 제어할 수 있는 이론적 근거를 마련하였다. 넷째, 실험을 통해 k -계수 이미지 매칭이 노이즈 제거 효과를 나타냄을 확인하고, 성능적인 측면에서 매우 우수함을 입증하였다. 실험 결과, 사용자는 계수 k 의 변경을 통해 노이즈 제거 정도를 제어할 수 있음을 확인하였다. 또한, 제안한 인덱스 기반 방법은 순차 스캔에 비해 수 배에서 수십 배까지 매칭 성능을 크게 향상시킨 것으로 나타났다.

이 같은 결과를 볼 때, 제안한 k -계수 이미지 매칭과 이의 인덱스 기반 해결책은 노이즈에 강건한 윤곽선 이미지 매칭을 수행할 수 있는 매우 실용적인 방법이라 생각한다. 특히, 본 연구는 이미지 도메인의 문제를 시계열 도메인의 전처리 변환을 통해 해결한 첫 번째 시도로서, 추후 스케일링, 회전 등의 다른 왜곡에도 활용될 수 있는 우수한 연구 결과라 사료된다. 본 논문의 향후 연구로는 시계열 길이에 따른 노이즈 제거 효과에 대한 비교 연구와 제안한 기법을 기반으로 텍스처 및 컬러 등의 특징을 결합한 멀티모달 이미지 매칭 기법에 대해 연구를 진행할 예정이다.

참고 문헌

- [1] Agrawal, R., Faloutsos, C., and Swami, A., "Efficient Similarity Search in Sequence Databases," In *Proc. the 4th Int'l Conf. on Foundations of Data Organization and Algorithms*, Chicago, Illinois, pp. 69-84, Oct. 1993.
- [2] Faloutsos, C., Ranganathan, M., and Manolopoulos, Y., "Fast Subsequence Matching in Time-Series Databases," In *Proc. Int'l Conf. on Management of Data*, ACM SIGMOD, Minneapolis, Minnesota, pp.419-429, May 1994.
- [3] Keogh, E., Wei, L., Xi, X., Lee, S.-H., and Vlachos, M., "LB_Keogh Supports Exact Indexing of Shapes under Rotation Invariance with Arbitrary Representations and Distance Measures," In *Proc. the 32rd Int'l Conf. on Very Large Data Bases*, Seoul, Korea, pp.882-893, Sept. 2006.
- [4] Moon, Y.-S., Whang, K.-Y., and Han, W.-S., "General Match: A Subsequence Matching Method in Time-Series Databases Based on Generalized Windows," In *Proc. Int'l Conf. on Management of Data*, ACM SIGMOD, Madison, Wisconsin, pp. 382-393, June 2002.
- [5] Loh, W.-K., Park, Y.-H., and Yoon, Y.-I., "Fast

- Recognition of Asian Characters Based on Database Methodologies," In *Proc. the 24th British Nat'l Conf. on Databases*, Glasgow, UK, pp.37-48, July 2007.
- [6] Vlachos, M., Vagena, Z., Yu, P. S., and Athitsos, V., "Rotation Invariant Indexing of Shapes and Line Drawings," In *Proc. of ACM Conf. on Information and Knowledge Management*, Bremen, Germany, pp.131-138, Oct. 2005.
- [7] Lee, A. J. T. et al., "A Novel Filtration Method in Biological Sequence Databases," *Pattern Recognition Letters*, vol.28, Issue 4, pp.447-458, Mar. 2007.
- [8] Brailean, J. C. et al., "Noise Reduction Filters for Dynamic Image Sequences: A Review," In *Proceedings of the IEEE*, vol.83, no.9, Sept. 1995.
- [9] Rosenfeld, A. and Kalk, A. C, *Digital Picture Processing*, Vol. 1/2, 2nd Ed., Academic Press, New York, 1982.
- [10] Moon, Y.-S. and Kim, J., "Efficient Moving Average Transform-Based Subsequence Matching Algorithms in Time-Series Databases," *Information Sciences*, vol.177, no.23, pp.5415-5431, Dec. 2007.
- [11] Rafiei, D. and Mendelzon, A. O., "Querying Time Series Data Based on Similarity," *IEEE Trans. on Knowledge and Data Engineering*, vol.12, no.5, pp.675-693, Sept. 2000.
- [12] Keogh, E., "Exact Indexing of Dynamic Time Warping," In *Proc. the 28th Int'l Conf. on Very Large Data Bases*, Hong Kong, pp.406-417, Aug. 2002.
- [13] Han, W.-S., Lee, J., Moon, Y.-S., and Jiang, H., "Ranked Subsequence Matching in Time-Series Databases," In *Proc. the 33rd Int'l Conf. on Very Large Data Bases*, Vienna, Austria, pp.423-434, Sept. 2007.
- [14] Loh, W.-K., Kim, S.-W., and Whang, K.-Y., "Index Interpolation: A Subsequence Matching Algorithm Supporting Moving Average Transform of Arbitrary Order in Time-Series Databases," *IEICE Transactions on Information and Systems*, vol.E84-D, no.1, pp.76-86, 2000.
- [15] Chatfield, C., *The Analysis of Time Series: An Introduction*, 3rd Ed., Chapman and Hall, 1984.
- [16] Beckmann, N., Kriegel, H.-P., Schneider, R., and Seeger, B., "The R*-tree: An Efficient and Robust Access Method for Points and Rectangles," In *Proc. Int'l Conf. on Management of Data*, ACM SIGMOD, Atlantic City, New Jersey, pp.322-331, May 1990.
- [17] Keogh, E. and Ratanamahatana, C. A., "Indexing and Mining Large Time Series Databases," In *Proc. The 12th Int'l Conf. on Database Systems for Advanced Applications*, Tutorial, Bangkok, Thailand, Apr. 2007.
- [18] Gonzalez, R. C. and Woods, R. E., *Digital Image Processing*, 2nd Ed., Prentice Hall, New Jersey, 2002.
- [19] Theoharatos, C., "A Generic Scheme for Color Image Retrieval Based on the Multivariate Wald-Wolfowitz Test," *IEEE Trans. on Knowledge and Data Engineering*, vol.17, no.6, pp.808-819, June 2005.
- [20] Do, M. N., "Wavelet-Based Texture Retrieval Using Generalized Gaussian Density and Kullback-Leibler Distance," *IEEE Trans. on Image Processing*, vol.11, no.2, Feb. 2002.
- [21] Zhang, D. Z. and Lu, G., "Review of Shape Representation and Description Techniques," *Pattern Recognition*, vol.37, no.1, pp.1-19, July 2003.



김범수

2006년 2월 한라대학교 컴퓨터공학과 학사. 2008년 2월 강원대학교 컴퓨터과학과 석사. 2008년 3월~현재 강원대학교 컴퓨터과학과 박사과정. 관심분야는 Data Mining & Knowledge Discovery, Data Warehousing & OLAP, Data Mining Applications(Images, Bioinformatics, Multimedia, Spatio-Temporal Databases, etc.)



문양세

1991년 2월 한국과학기술원 과학기술대학 전산학과 학사. 1993년 2월 한국과학기술원 전산학과 석사. 2001년 8월 한국과학기술원 전자전산학과 전산학전공 박사. 1993년 2월~1997년 2월 현대전자산업(주) 주임연구원. 2001년 9월~2002년 2월 (주)현대시스템 선임연구원 2002년 2월~2005년 2월 (주)인프라밸리 기술위원(이사). 2005년 3월~2008년 2월 한국과학기술원 첨단정보기술연구소 연구원. 2008년 7월~2009년 8월 미국 퍼듀대학교 방문연구원. 2005년 3월~현재 강원대학교 컴퓨터과학과 부교수. 관심분야는 Data Mining, Knowledge Discovery, Stream Data, Storage System, Database Applications, Mobile/ Wireless Communication Services & Systems



김진호

1982년 2월 경북대학교 전자공학과 학사 1985년 2월 한국과학기술원 전산학과 석사. 1990년 2월 한국과학기술원 전산학과 박사. 1995년 8월~1996년 7월 미국 미시간 대학교 객원 교수. 2003년 2월~2004년 2월 미국 Drexel University 객원 교수. 1999년 3월~2008년 2월 한국과학기술원 첨단정보기술연구소 연구원. 1990년 8월~현재 강원대학교 컴퓨터과학과 교수. 관심분야는 Data warehouse, OLAP, Data Mining, Real-time/Embedded Database, Main-memory database, Data Modeling, Web Database Technology