DOI: 10.3745/KIPSTD.2011.18D.1.009

# 엔빌로프 기반 하한을 사용한 효율적인 회전-불변 윤곽선 이미지 매칭

김 상 필[†] · 문 양 세[††] · 홍 선 경[†††]


## 요 약

본 논문에서는 윤곽선 이미지 매칭에서 회전-불변 거리 계산의 효율적 방법을 제안한다. 회전-불변 거리 계산은 이미지 시계열을 한 칸씩 회전하면서 매번 유클리디안 거리를 계산해야 하는 고비용의 연산이다. 본 논문에서는 엔빌로프 기반 하한을 사용하여 회전-불변 거리 계산을 크게 줄이는 획기적인 해결책을 제시한다. 이를 위해, 먼저 질의 시퀀스 대상의 단일 엔빌로프 작성과 이의 하한 개념을 제시하고, 이를 회전-불변 거리 계산에 사용하면 많은 수의 회전-불변 거리 계산을 줄일 수 있음을 보인다. 그런데, 단일 엔빌로프 기법은 하나의 엔빌로프가 가능한 모든 회전 시퀀스를 포함하기 때문에 하한이 커지고, 이에 따라 매칭 성능이 저하되는 문제점이 있다. 이러한 문제점을 해결하기 위하여, 본 논문에서는 회전 구간의 개념을 도입하여 단일 엔빌로프 기반 하한을 다중 엔빌로프 기반 하한 개념으로 확장한다. 또한, 다중 엔빌로프 기법에서 회전 구간을 결정하기 위한 방법으로 동일-너비 기법과 엔빌로프 최소화 기법을 제안한다. 실험 결과, 제안한 엔빌로프 기반 매칭 기법은 기존 기법에 비해 최대 수 배에서 수십 배까지 매칭 성능을 향상시킨 것으로 나타났다.


**키워드 : 윤곽선 이미지 매칭, 데이터 마이닝, 회전-불변 거리, 유사 시퀀스 매칭, 엔빌로프 기반 하한**

# Efficient Rotation-Invariant Boundary Image Matching Using the Envelope-based Lower Bound

Sang-Pil Kim[†] · Yang-Sae Moon[††] · Sun-Kyong Hong[†††]


## ABSTRACT

In this paper we present an efficient solution to rotation invariant boundary image matching. Computing the rotation-invariant distance between image time-series is a time-consuming process since it requires a lot of Euclidean distance computations for all possible rotations. In this paper we propose a novel solution that significantly reduces the number of distance computations using the envelope-based lower bound. To this end, we first present how to construct a single envelope from a query sequence and how to obtain a lower bound of the rotation-invariant distance using the envelope. We then show that the single envelope-based lower bound can reduce a number of distance computations. This approach, however, may cause bad performance since it may incur a larger lower bound by considering all possible rotated sequences in a single envelope. To solve this problem, we present a concept of rotation interval, and using the rotation interval we generalize the envelope-based lower bound by exploiting multiple envelopes rather than a single envelope. We also propose equi-width and envelope minimization divisions as the method of determining rotation intervals in the multiple envelope approach. Experimental results show that our envelope-based solutions outperform existing solutions by one or two orders of magnitude.


**Keywords : Boundary Image Matching, Data Mining, Rotation-Invariant Distance, Similar Sequence Matching, Envelope-Based Lower Bound**

※ 이 논문은 2010년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(2010-0002518).
† 준 회 원 : 강원대학교 컴퓨터과학과 석사과정
†† 종신회원 : 강원대학교 컴퓨터과학과 부교수(교신저자)
††† 준 회 원 : 강원대학교 컴퓨터과학과 박사과정
논문접수 : 2010년 10월 15일
수 정 일 : 1차 2010년 11월 23일
심사완료 : 2010년 11월 23일

## 1. 서 론

컴퓨터 계산 및 저장 능력의 발전에 따라, 대용량 시계열 데이터베이스 대상의 시계열 매칭(time-series matching) 연구가 활발하게 이루어져 왔다[1, 2, 3, 4]. 또한, 최근에는 필기체 인식[5], 이미지 매칭[6, 7], 바이오 시퀀스 매칭[8] 등 여러 응용에서 시계열 매칭 연구가 활용되고 있다. 본 논문에서는 이중 대용량 이미지 데이터베이스를 대상으로 하는 윤곽선 이미지 매칭 문제를 다룬다. 윤곽선 이미지 매칭은 (그림 1)과 같이 이미지의 윤곽선을 시계열로 변환한 후, 시계열 매칭을 사용하여 유사 이미지를 찾는 방법이다[6, 7, 9].

윤곽선 이미지 매칭의 최근 연구 중 주목 받는 내용이 회전-불변(rotation-invariance)의 지원이며[6, 9], 두 이미지 시퀀스의 회전-불변 거리는 다음과 같이 정의한다.

**[정의 1]** 길이 $n$인 두 시퀀스 $Q(=\{q_0,\ldots,q_{n-1}\})$와 $S(=\{s_0,\ldots,s_{n-1}\})$의 회전 불변 거리 $RID(Q,S)$는 다음 식 (1)과 같이 정의한다.

$$RID(Q,S)=\min_{j=0}^{n-1} D(Q^j,S)=\min_{j=0}^{n-1}\sqrt{\sum_{i=0}^{n-1}\left|q_{(i+j)\%n}-s_i\right|^2} \quad (1)$$

식 (1)에서 $D(Q,S)$는 두 시퀀스 $Q$와 $S$의 유클리디안 거리인 $\sqrt{\sum_{i=0}^{n-1}|q_i-s_i|^2}$을 나타낸다. ■

[정의 1]에서 $Q^j$는 시퀀스 $Q$를 j만큼 회전하여 얻은 시퀀스 $\{q_j,\ldots,q_{n-1},q_0,\ldots,q_{j-1}\}$를 나타내며, 본 논문에서는 $Q^j$를 Q의 $j$-회전 시퀀스라 부른다. 회전 불변 거리는 모든 가능한 $j$-회전 시퀀스 $Q^j$를 고려했을 때 얻을 수 있는 최소 거리로서, 식 (1)의 계산을 위해서는 $\Theta(n)$번의 유클리디안 거리 계산, 즉 $\Theta(n^2)$의 계산 복잡도가 요구된다.

**[정의 2][1)]** 질의 시퀀스 Q와 허용치(tolerance) ε이 주어졌을 때, Q와의 회전 불변 거리가 ε 이하인 모든 데이터 시퀀스를 찾는 작업을 회전-불변 (윤곽선) 이미지 매칭이라 한다. ■

1) [정의 2]의 범위 질의(range query) 외에 $k$-최근접($k$-NN: $k$-nearest neighbor) 질의도 많이 사용된다. 그런데, $k$-NN 질의도 현재까지 구한 $k$개 후보의 거리를 허용치로 사용하는 범위 질의로 해석할 수 있다. 이에 따라, 본 논문에서는 허용치가 주어지는 범위 질의에 초점을 맞추어 연구를 진행한다.

이와 같은 회전-불변 이미지 매칭에서는 길이 n인 모든 데이터 시퀀스에 대해 $\Theta(n)$번의 많은 유클리디안 거리 계산이 필요하고, 이는 성능 저하의 주된 요인이 된다[6, 9].

본 논문에서는 엔빌로프(envelope) 기반의 하한을 사용하여 윤곽선 이미지 매칭에서 빈번하게 계산되는 회전 불변 거리 계산 횟수를 획기적으로 줄이는 방법을 제안한다. 이를 위해, 우선 단일 엔빌로프 개념을 제안한다. 질의 시퀀스 $Q$의 엔빌로프 $[L,U]$는 $Q$의 모든 가능한 $j$-회전 시퀀스를 포함하는 고차원 MBR(minimum bounding rectangle)로서, $L$은 최소값 엔트리들로 구성된 시퀀스를, $U$는 최대값 엔트리들로 구성된 시퀀스를 각각 나타낸다. 본 논문에서는 $Q$의 엔빌로프 $[L,U]$와 데이터 시퀀스 $S$간의 거리 $(=D([L,U],S))$가 $Q$와 $S$간 회전-불변 거리 하한임을 증명하고, 이를 회전-불변 이미지 매칭에 활용한다. 하한 $D([L,U],S)$가 주어진 허용치 이상일 경우, 실제 회전-불변 거리는 계산 필요가 없는 성질을 활용한다. 본 논문에서는 단일 엔빌로프 기반의 하한을 사용하여, 회전-불변 거리 계산 횟수를 크게 줄인 회전-불변 이미지 매칭 알고리즘을 제안한다.

그런데, 단일 엔빌로프 기반 매칭 알고리즘은 하한이 작아서 전지(pruning) 효율이 좋지 않은 문제점이 있다. 이는 단일 엔빌로프 $[L,U]$가 모든 회전 시퀀스를 고려하여 그 넓이가 넓어지고, 결국 하한이 작아지기 때문이다. 이 문제점을 해결하기 위하여, 본 논문에서는 회전 구간 개념을 도입하여 단일 엔빌로프 기반 하한을 다중 엔빌로프 기반 하한 개념으로 확장한다. 회전 구간은 질의 시퀀스 $Q$를 서로소(disjoint)인 여러 구간으로 나눈 것을 말하며, 다중 엔빌로프는 각 회전 구간에 해당하는 회전 시퀀스만을 고려하여 엔빌로프들을 구성한다. 이러한 다중 엔빌로프를 사용하면 각 엔빌로프의 넓이가 줄어 들어 하한이 커지게 되고, 결국 전지 효과가 크게 발휘된다. 본 논문에서는 다중 엔빌로프 기반 하한을 정형적으로 제시하고, 이를 기반으로 다중 엔빌로프 기반 매칭 알고리즘을 제안한다. 더 나아가 회전 구간의 크기를 결정하는 방법으로 (1) 구간 크기를 동일하게 나누는 동일-너비 기법과 (2) 엔빌로프들의 넓이를 최소화하고자 하는 엔빌로프 최소화 기법을 제시한다. 실험 결과 제안한 엔빌로프 기반 알고리즘은 기존 알고리즘에 비해 불필요한 회전-불변 거리 계산을 획기적으로 줄이고, 이를 통해 성능을 크게 향상시킨 것으로 나타났다. 특히, 다중 엔빌로프 기반 알고리즘은 기존 알고리즘에 비해 회전-불변 거리 계산 횟수는 수십 배까지 줄이고, 성능은 수 배까지 향상시킨 것으로 나타났다.

본 논문의 구성은 다음과 같다. 제2장에서는 회전-불변 이미지 매칭의 기존 연구를 설명한다. 제3장에서는 제안하

는 단일 엔빌로프 기반 하한을 사용한 이미지 매칭을 제시하고, 제4장에서는 다중 엔빌로프 기반 하한을 사용한 이미지 매칭을 설명한다. 제5장에서는 다중 엔빌로프 회전 구간 크기를 결정하는 방법을 제안한다. 제6장에서는 제안한 방법과 기존의 연구 결과를 비교한 성능 평가 결과를 제시하고, 마지막으로 제7장에서는 결론을 기술한다.

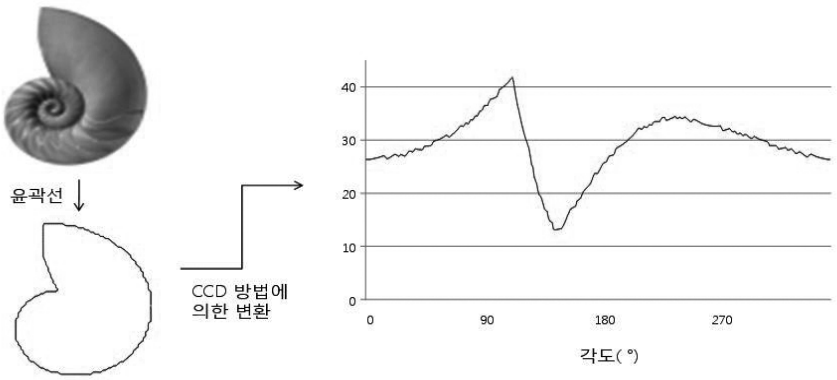
## 2. 관련 연구 및 기존 알고리즘 (Related Work and Previous Algorithms)

주어진 질의 시퀀스와 유사한 데이터 시퀀스를 찾는 과정인 시계열 매칭은 Agrawal 등[1]의 전체 매칭과 Faloutsos 등[2]의 서브시퀀스 매칭에서 시작하여 최근까지 많은 연구가 진행되었다. 시계열 매칭에서 사용하는 유사성 척도로는 유클리디안 거리[1, 4]와 DTW(dynamic time warping) 거리[3, 10]가 주로 사용되었다. 질의 종류로는 범위 질의[2, 4, 7]와 $k$-NN(nearest neighbor) 검색[10, 11]에 대한 연구가 진행되었다. 본 논문에서 다루는 윤곽선 이미지 매칭은 이러한 시계열 매칭의 중요한 응용으로 볼 수 있다.

이미지 매칭[12, 13]은 주어진 이미지와 유사한 이미지를 찾는 문제로 이미지 처리 분야에서의 주요 연구 분야 중 하나이다. 이미지 매칭을 위해 지금까지 다양한 이미지 정보를 이용하려는 시도들이 있었다. 예를 들어, 참고문헌 [14]에서는 색상을, 참고문헌 [15]에서는 질감을, 참고문헌 [16]에서는 모양을 이미지 매칭에 각각 이용하였다. 본 논문에서는 모양 기반의 이미지 매칭에 연구의 초점을 맞춘다. 모양 기반의 이미지 매칭에서는 주로 객체의 외부 윤곽선이나 영역을 이용한다[17]. 객체의 외부 윤곽선 추출 방법에는 Chain Code[18], Polygon[19], CSS(curvature scale space)[20], CCD(centroid contour distance)[6, 7, 9, 16] 등이 있다. 본 논문에서는 외부 윤곽선을 이용하는 간단한 방법인 CCD 방법을 활용한다. 이외의 방법을 이용한 회전-불변 윤곽선 이미지 매칭 연구와 이들 간의 비교는 향후 연구로 다룬다. (그림 1)에서 보듯이, CCD 방법은 이미지의 외부 윤곽선 중심점을 찾은 후, y축(0°)에서 시작하여 일정한 각도($\Delta\theta = 2\pi/n$)의 시계 방향(혹은 시계 반대 방향)으로 진행하며 중심점과 외부 윤곽선과의 거리를 계산하여, 이미지를 n-차원 공간의 점으로 매핑하는 방법이다. 이와 같이 CCD 방법을 사용하면 외부 윤곽선 이미지를 시계열로 나타낼 수 있고, 이에 따라서 시계열 매칭 기법을 이미지 매칭에 활용할 수 있다[6, 9].

회전-불변 윤곽선 이미지 매칭에 관한 최근 연구는 다음과 같다. 먼저, Vlachos 등[6]은 인덱스를 사용하여 성능을 향상시키는 방법을 제안하였다. 이 연구에서는 DFT의 진폭이 회전-불변의 특성을 가짐을 보이고, 이를 인덱스 구축 및 필터링에 사용하였다. Keogh 등[9]은 회전-불변 이미지 매칭에 LB_Keogh[3]가 적용됨을 보이고, 이를 사용하여 회전-불변 거리 계산이 필요한 후보 개수를 크게 줄이는 방법을 제안하였다. 서론에서 언급한 바와 같이, 이들 연구는 필터링을 통해 회전-불변 거리 계산이 필요한 후보 데이터 시퀀스 개수를 줄이는 것이 목적이다. 따라서, 이들 방법들도 궁극적으로는 (후보) 데이터 시퀀스와 질의 시퀀스와의 회전-불변 거리 계산은 필요하며, 본 논문의 연구 결과는 이들 계산에 적용이 가능하므로, 기존 연구와 본 연구는 직교적이라 할 수 있다.

회전-불변 이미지 매칭의 기본 알고리즘 *RI-Naïve*는 (그림 2)과 같다. 그림에서 보듯이, *RI-Naïve*는 각 (후보) 데이터 시퀀스 $S$에 대해서, 질의 시퀀스와의 회전-불변 거리를 계산하고, 그 거리가 주어진 허용치 이하인지 판단한다. 다음으로, *RI-Naïve*에 미리 버림(early abandon)[9]을 적용한 알고리즘을 *RI-EA*라 한다. 미리 버림이란 유클리디안 거리 계산 과정에서 중간까지의 거리 값이 주어진 허용치보다 커지면 계산을 중단하는 방법이다. 즉, 알고리즘 *RI-Naïve*의라인 4에서 유클리디안 거리 $D(Q^j, S)$를 계산하는 과정 내부에서, 현재까지의 거리 제곱의 합이 허용치 제곱보다 크면 거리 계산을 중단하는 구조를 갖는다. 미리 버림을 사용하는 것을 제외하고는 *RI-EA*와 *RI-Naïve*는 동일한 구조를 갖는




(그림 1) 이미지의 윤곽선 추출 및 이의 시계열 변환 예제

다. 기존 연구[6, 9]에서는 후보 데이터 시퀀스들을 구한 후에, *RI-Naïve*나 *RI-EA*에 해당하는 매칭을 수행해야 하며, 본 논문에서는 이러한 매칭의 성능을 크게 향상시킨다.

```
RI-Naïve(query sequence Q, a set S of data sequences, tolerance ε)
1.  R := ∅ ;
2.  for each data sequence S ∈ S do
3.      for j := 0 to (n − 1) do
4.          if D(Q^j, S) ≤ ε then
5.              R := R ∪ {S} ;
6.              break;
7.          end-if
8.      end-for
9.  end-for
10. return R ;
```

(그림 2) 회전-불변 이미지 매칭의 기본 알고리즘 *RI-Naïve*

## 3. 단일 엔빌로프 기반의 하한과 매칭 알고리즘

*RI-Naïve*와 *RI-EA*의 문제점은 회전-불변 거리 계산의 횟수가 많다는 점이다. 길이 $n$ 인 두 시퀀스에 대한 *RI-Naïve*와 *RI-EA*의 회전-불변 거리 계산은 $\Theta(n^2)$ 의 복잡도를 가지는데, 이는 한 시퀀스를 고정하고 다른 시퀀스를 $n$ 번 회전해 가며 거리를 계산하기 때문이다. 특히, 비교해야 하는 데이터 시퀀스가 많은 경우, 이러한 계산 복잡도는 매칭 성능을 저하시키는 주요 원인이 된다. 따라서, 본 논문에서는 엔빌로프 기반의 회전 불변 거리 하한을 제시하고, 이를 윤곽선 이미지 매칭에 활용하는 방법을 제안한다. 엔빌로프 기반의 하한은 질의 시퀀스의 엔빌로프와 데이터 시퀀스 간의 거리로 계산된다.

**[정의 3]** 길이가 $n$ 인 질의 시퀀스 $Q$ 가 주어졌을 때, 같은 길이의 두 시퀀스 $L$ 과 $U$ 의 각 엔트리 $l_i$ 와 $u_i$ 는 다음 식 (2)와 같이 계산하고, 이들 시퀀스 $[L,U]$ 를 질의 시퀀스 $Q$ 의 엔빌로프라고 한다. 또한, 질의 시퀀스 $Q$ 의 엔빌로프 $[L,U]$ 와 데이터 시퀀스 $S$ 와의 거리 $D([L,U],S)$ 는 다음 식 (3)과 같이 계산한다.

$$l_i = \min_{j=0}^{n-1} q_{(i+j)\%n},\ u_i = \max_{j=0}^{n-1} q_{(i+j)\%n} \tag{2}$$

$$D([L,U],S) = \sqrt{\sum_{i=0}^{n-1} \begin{cases} |s_i - u_i|^2 & \text{if } s_i > u_i; \\ |s_i - l_i|^2 & \text{if } s_i < l_i; \\ 0 & \text{otherwise.} \end{cases}} \tag{3}$$

■

[보조 정리 1]은 질의 시퀀스 $Q$ 의 엔빌로프 $[L,U]$ 와 데이터 시퀀스 $S$ 와의 거리 $D([L,U],S)$ 가 $Q$ 와 $S$ 의 회전-불변 거리의 하한임을 나타낸다.

**[보조정리 1]** 질의 시퀀스 $Q$ 의 엔빌로프 $[L,U]$ 와 데이터 시퀀스 $S$ 와의 거리 $D([L,U],S)$ 는 $Q$ 와 $S$ 의 회전-불변 거리 $RID(Q,S)$ 의 하한이다.

**[증 명]** 두 시퀀스 $Q$ 와 $S$ 의 유클리디안 거리 $D(Q,S)$ 는 $\sqrt{\sum_{i=0}^{n-1}|q_i - s_i|^2}$ 이다. 시계열 $L$ 의 엔트리는 질의 시퀀스의 가장 작은 값으로 만들어진 시퀀스이고, 시계열 $U$ 의 엔트리는 질의 시퀀스의 가장 큰 값으로 만들어진 시퀀스로서, 질의 시퀀스 $Q$ 의 모든 엔트리는 $L$ 과 $U$ 사이에 존재하게 된다. 즉, $l_i \le q_i \le u_i$ 의 관계가 성립한다. 여기서, 만일 $s_i > u_i$ 라면, $q_i \le u_i$ 에 의해 $|s_i - q_i| \ge |s_i - u_i|$ 가 성립하고, $s_i < l_i$ 라면, $l_i \le q_i$ 에 의해 $|s_i - q_i| \ge |s_i - l_i|$ 가 성립하며, 그렇지 않을 경우 $(l_i \le s_i \le u_i)$ 는 $|s_i > q_i| \ge 0$ 가 당연히 성립한다. 따라서, $(s_i - u_i)^2$, $(s_i - l_i)^2$, 0 을 더해 계산하는 $D([L,U],S)$ 는 $(s_i - q_i)^2$ 을 더해 계산하는 $RID(Q,S)$ 이하이므로, $D([L,U],S)$ 는 $RID(Q,S)$ 의 하한이다. ■

본 논문에서는 [정의 3]의 하한 $D([L,U],S)$ 를 모든 가능한 회전 시퀀스를 고려해 "하나의 엔빌로프"를 구성한다는 의미에서 $LB_{SE}(Q,S)$ 로 표기한다. (SE는 single envelope를 의미함.) 하한 $LB_{SE}(Q,S)$ 를 사용하면, 실제 회전-불변 거리를 계산하지 않고도 유사하지 않은 많은 데이터 시퀀스를 미리 전지할 수 있다. 즉, $LB_{SE}(Q,S)$ 가 주어진 허용치보다 크면, 실제 회전-불변 거리인 $RID(Q,S)$ 를 계산할 필요 없이 유사하지 않은 것으로 판별할 수 있다. 결국, $LB_{SE}(Q,S)$ 가 허용치보다 작은 경우에만 $RID(Q,S)$ 를 계산함으로써, 회전 불변 거리를 계산하는 횟수를 크게 줄일 수 있다. (그림 3)은 $LB_{SE}(Q,S)$ 를 사용한 회전-불변 윤곽선 이미지 매칭 알고리즘 $RI-SE$ 를 나타낸다. $LB_{SE}(Q,S)$ 의 계산 복잡도는 $\Theta(n)$ 으로 $RID(Q,S)$ 의 $\Theta(n^2)$ 에 비해 낮고, 따라서 전체적인 매칭 성능을 향상시킬 수 있다.

```
Algorithm RI-SE(query sequence Q, a set S of data sequences, tolerance ε)
1.  Construct L and U from Q;
2.  R := ∅;
3.  for each data sequence S∈S do
4.      if LB_SE(Q,S) ≤ ε then
5.          if RID(Q,S) ≤ ε then
6.              R := R ∪ {S};
7.          end-if
8.      end-if
9.  end-for
10. return R;
```

(그림 3) 엔빌로프 기반의 하한을 이용한 회전 불변 이미지 매칭 알고리즘

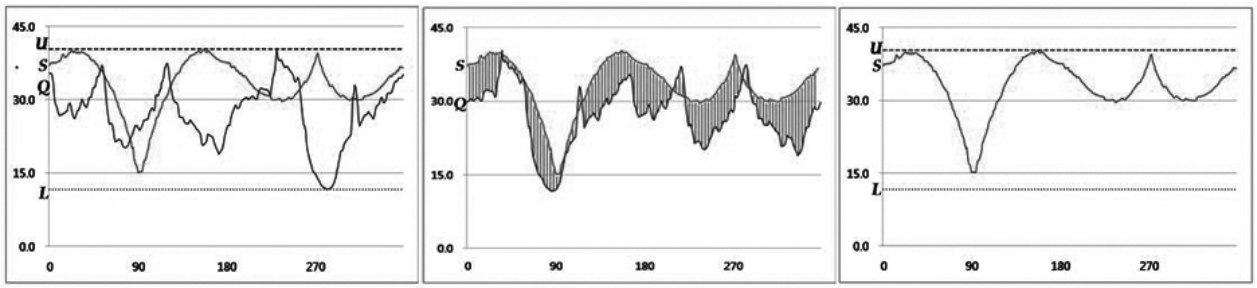
알고리즘 $RI-SE$ 의 문제점은 질의 시퀀스 $Q$ 의 엔트리 값 변화가 큰 경우, 엔빌로프가 넓어져(즉, $LB_{SE}(Q,S)$ 는 작아져) 전지 효과가 크게 발휘되지 않을 수 있다는 것이다. 다음 예제 1이 이러한 예를 설명한다.

**[예제 1]** (그림 4)는 윤곽선 이미지를 길이 360의 시퀀스로 표현한 것이다. (그림 4(a))는 질의 시퀀스 $Q$ 와 이의 엔빌로프 $[L,U]$, 그리고 데이터 시퀀스 $S$ 를 나타낸다. (그림 4(b))의 빗금친 부분은 회전 불변 거리 $RID(Q,S)$ 를, (그림 4(c))는 이의 하한인 $LB_{SE}(Q,S)$ 를 각각 나타낸다. 그림에서 보듯이, 엔빌로프를 구성하는 $L$ 의 모든 엔트리는 $Q$ 의 엔트리 중 최소값으로 결정되며, $U$ 의 모든 엔트리는 $Q$ 의 엔트리 중 최대값으로 결정된다. 즉, 엔빌로프 $[L,U]$ 가 너무 넓게 결정되어, 이를 사용하여 계산되는 하한 $LB_{SE}(Q,S)$ 는 회전-불변 거리인 $RID(Q,S)$ 에 비해 너무 작은 값을 가지게 된다. (그림 4(b))의 경우 $RID(Q,S)$ = 137.8인 반면, (그림 4(c))의 하한 $LB_{SE}(Q,S)$ = 0으로 계산되어, 하한이 제대로 역할을 하지 못한다. 이와 같이 하한이 작을 경우, 매칭 과정에서 미리 전지하는 효과가 크게 발휘되지 않는 문제점이 발생한다. ■

(그림 5)는 [예제 1](그림 4)의 문제점을 이미지 도메인에서 나타낸 것이다. (그림 5(a))에서 나뭇잎 윤곽선은 질의 시퀀스 $Q$ 를, 바깥쪽 원은 엔빌로프의 $U$ 를, 안쪽 원은 $L$ 을 각각 나타낸다. 그리고, 하트 윤곽선은 데이터 시퀀스 $S$ 를 나타낸다. (그림 5(b))의 빗금 친 부분은 회전 불변 거리 $RID(Q,S)$ 가 의미하는 바를, (그림 5(c))는 $RID(Q,S)$ 의 하한인 $LB_{SE}(Q,S)$ 가 의미하는 바를 각각 나타낸다. 그림에서 볼 수 있듯이, 질의 이미지에서 생성되는 엔빌로프의 최대값 및 최소값의 차이가 클 경우, 즉 엔빌로프가 넓을수록 $LB_{SE}(Q,S)$ 는 작은 값을 가지게 된다. (그림 5)의 경우 하트 이미지 데이터 시퀀스가 엔빌로프 안에 완전히 포함되어 $LB_{SE}(Q,S)$ 의 값이 0으로 계산되고, 하한으로서 제 역할을 하지 못하게 된다. 이는, 질의로 주어지는 이미지 윤곽선의 변화가 클수록 $LB_{SE}(Q,S)$ 는 작은 값을 갖게 되고 결국 전지 효과가 발휘되지 못함을 의미한다.

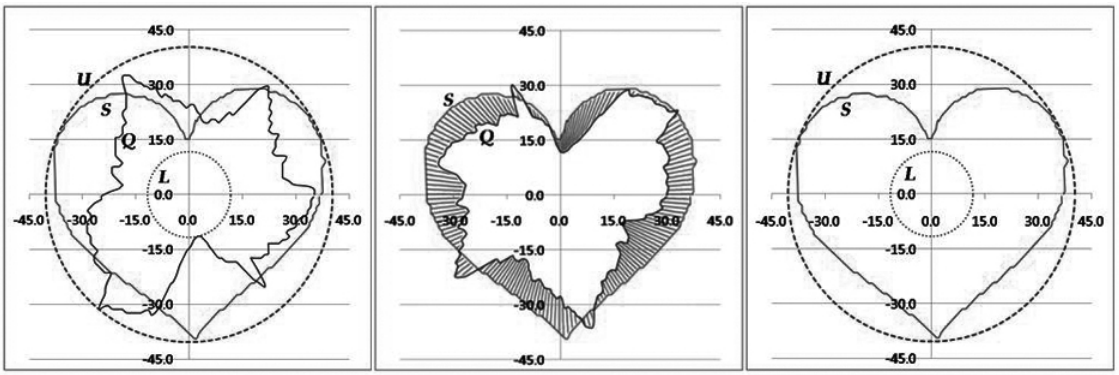
## 4. 다중 엔빌로프 기반 하한과 매칭 알고리즘

단일 엔빌로프에서 하한이 커지는 문제점을 해결하기 위하여, 본 절에서는 엔빌로프를 여러 개 사용하는 다중 엔빌로프 기반 하한의 개념을 제시한다. 단일 엔빌로프 접근법은 질의 시퀀스의 모든 가능한 회전 시퀀스를 고려하여 엔빌로프를 구하기 때문에, 그 하한 값이 커지는 문제점이 발생한다. 좀 더 자세히 설명하면, 엔빌로프 기반 하한 $LB_{SE}(Q,S)$ 는 길이가 $n$ 인 질의 시퀀스 $Q$ 를 모든 가능한 $j$ ($=0,...,n-1$)에 대하여 회전한 $Q^j$ 를 고려하기 때문에




(a) 시퀀스와 엔빌로프. (b) 회전-불변 거리 $RID(Q,S)$. (c) 엔빌로프 기반 하한 $LB_{SE}(Q,S)$.

(그림 4) 질의 시퀀스 $Q$ 와 데이터 시퀀스 $S$ 의 회전 불변 거리 및 이의 하한




(a) 윤곽선 이미지와 엔빌로프. (b) 회전-불변 거리 $RID(Q,S)$. (c) 엔빌로프 기반 하한 $LB_{SE}(Q,S)$.

(그림 5) 이미지 도메인에서 본 질의 시퀀스 $Q$와 데이터 시퀀스 $S$의 회전 불변 거리 및 이의 하한

그 값이 너무 커지는 문제가 생긴다. 따라서, 본 절에서는 가능한 $j$ 값을 여러 구간으로 나누고 각 구간에 대해 하한을 구한 후, 이 하한의 최소값을 전체 구간에 대한 하한으로 삼는 방법을 사용한다. 이는 고려하는 $j$ 값 구간이 작을수록 해당 구간에서의 하한은 커지게 되고, 이에 따라 전체 구간의 하한 또한 $LB_{SE}(Q,S)$ 보다는 커지게 될 것이라는 직관에 기반한다. 이 방법을 설명하기 위해 우선 회전 불변 거리 개념을 '구간'을 사용하여 일반화한다.

**[정의 4]** 길이 $n$인 질의 시퀀스 $Q$를 $a, a+1, ..., b$ 만큼씩 회전하여 얻은 시퀀스 $Q^a$, $Q^{a+1}$, …, $Q^b$와 데이터 시퀀스 $S$와 거리의 최소값을 구간 [a,b]에서의 $Q$와 $S$의 회전 불변 거리라 정의하고, $RID(Q^{[a,b]},S)$로 표기한다. 즉, $RID(Q^{[a,b]},S)$는 다음 식 (4)와 같이 계산한다.

$$RID\left(Q^{[a,b]},S\right)=\min_{j=a}^{b} D\left(Q^{j},S\right)=\min_{j=a}^{b}\sqrt{\sum_{i=0}^{n-1}\left|q_{(i+j)\%n}-s_i\right|^2} \qquad (4)$$

이때, 구간 [a,b]를 회전 불변 거리 $RID(Q^{[a,b]},S)$의 회전 구간이라 부른다. ■

다음으로 회전 구간 [a,b]를 사용한 엔빌로프 구성과 이의 하한 성질을 설명한다. 다음 [정의 5]는 질의 시퀀스에 대해 회전 구간을 고려한 엔빌로프 개념을 제시하고, 이 엔빌로프와 데이터 시퀀스 간의 거리를 정의한다.

**[정의 5]** 길이 $n$인 질의 시퀀스 $Q$와 회전 구간 [a,b]가 주어졌을 때, 두 시퀀스 $L^{[a,b]}, U^{[a,b]}$의 각 엔트리 $l_i^{[a,b]}$와 $u_i^{[a,b]}$는 식 (5)와 같이 계산되고, 이들 시퀀스 $[L^{[a,b]},U^{[a,b]}]$를 회전 구간 [a,b]에서 $Q$의 엔빌로프라 정의한다. 또한, 엔빌로프 $[L^{[a,b]},U^{[a,b]}]$와 데이터 시퀀스 $S$와의 거리 $D([L^{[a,b]},U^{[a,b]}],S)$는 다음 식 (6)과 같이 계산한다.

$$l_i^{[a,b]}=\min_{j=a}^{b} q_{(i+j)\%n}\ ,\quad u_i^{[a,b]}=\max_{j=a}^{b} q_{(i+j)\%n} \qquad (5)$$

$$D\left(\left[L^{[a,b]},U^{[a,b]}\right],S\right)=\sqrt{\sum_{i=0}^{n-1}\begin{cases}\left|s_i-u_i^{[a,b]}\right|^2 & \text{if } s_i>u_i^{[a,b]};\\ \left|s_i-l_i^{[a,b]}\right|^2 & \text{if } s_i<l_i^{[a,b]};\\ 0 & \text{otherwise.}\end{cases}} \qquad (6)$$

■

직관적으로 설명했을 때, 회전 구간 [a,b]에서 $Q$의 엔빌로프는 (모든 회전 시퀀스가 아닌) 회전 시퀀스 $Q^a, Q^{a+1}, ..., Q^b$ 만을 고려하여 엔빌로프를 구성한 것이다. [보조정리 2]는 엔빌로프 $[L^{[a,b]},U^{[a,b]}]$와 데이터 시퀀스 $S$와의 거리 $D([L^{[a,b]},U^{[a,b]}],S)$가 회전 구간 [a,b]에서의 $Q$와 $S$의 회전-불변 거리 $RID(Q^{[a,b]},S)$의 하한임을 나타낸다.

**[보조정리 2]** 회전 구간 [a,b]에서 질의 시퀀스 $Q$의 엔빌로프 $[L^{[a,b]},U^{[a,b]}]$와 데이터 시퀀스 $S$와의 거리 $D([L^{[a,b]},U^{[a,b]}],S)$는 $RID(Q^{[a,b]},S)$의 하한이다.

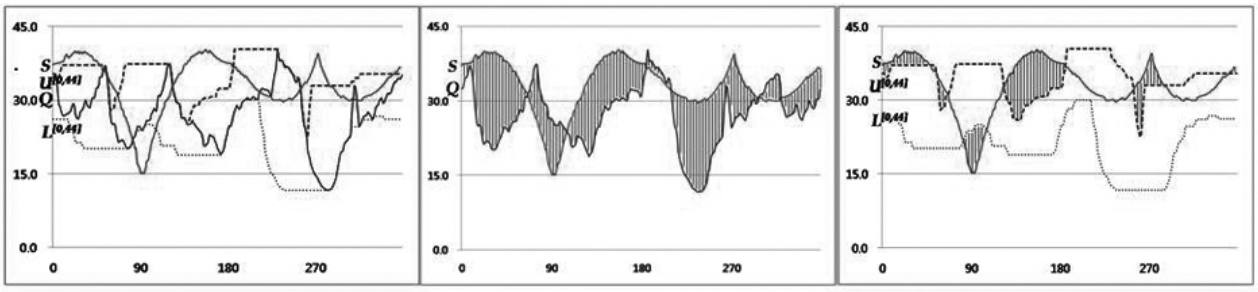
**[증 명]** [보조정리 1]과 동일하게 증명되므로, 자세한 과정을 생략한다. ■

엔빌로프 기반 하한을 계산하는데 있어서, 전체 구간[정의 3]이 아닌 회전 구간[정의 5]만을 고려하는 경우 그 하한 값이 커지는 장점이 있다. 본 논문에서는 전체 구간 [0,n-1]을 고려한 하한 $LB_{SE}(Q,S)$와 대비하기 위해, 회전 구간 [a,b]만을 고려한 하한 $D([L^{[a,b]},U^{[a,b]}],S)$을 $LB_{[a,b]}(Q,S)$로 나타낸다. 이 표기법으로 다시 설명하면, $LB_{SE}(Q,S)$는 모든 가능한 회전 시퀀스로 엔빌로프 [L,U]를 구성하고, 이 [L,U]로 계산한 하한인데 반해, $LB_{[a,b]}(Q,S)$는 회전 구간 [a,b] 내의 회전 시퀀스만으로 엔빌로프 $[L^{[a,b]},U^{[a,b]}]$를 구성하고, $[L^{[a,b]},U^{[a,b]}]$로 계산한 하한을 나타낸다. 이에 따라, 일반적으로 $[L^{[a,b]},U^{[a,b]}]$는 [L,U]보다 좁게 되고, 결과적으로 $LB_{[a,b]}(Q,S)$는 $LB_{SE}(Q,S)$보다 커지게 된다. 다음 [예제 2]가 이러한 예를 설명한다.

**[예제 2]** (그림 6)에서 질의 시퀀스 $Q$와 데이터 시퀀스 $S$는 (그림 4)와 동일한 시퀀스이며, 모든 시퀀스 길이는 360이다. (그림 4(a)와 4(c))의 엔빌로프는 회전 구간이 [0,359]로 모든 시퀀스를 고려한 반면에, (그림 6(a)와 6(c))의 엔빌로프는 회전 구간이 [0,44]로 45개의 시퀀스만을 고려한 것이다. (그림 6(a))는 이러한 엔빌로프 $[L^{[0,44]},U^{[0,44]}]$와 질의 시퀀스 $Q$, 데이터 시퀀스 $S$를 나타낸다. (그림 6(b))에서 빗금친 부분은 회전 불변 거리 $RID(Q^{[0,44]},S)$를, (그림 6(c))에서 빗금친 부분은 회전 구간 [a,b]만을 고려한 하한인 $LB_{[0,44]}(Q,S)$를 각각 나타낸다. (그림 6(a))의 엔빌로프 $[L^{[0,44]},U^{[0,44]}]$를 구성하는 $L^{[0,44]}$과 $U^{[0,44]}$의 각 엔트리는 질의 시퀀스 $Q$를 구성하는 $n$개 엔트리 모두를 고려하는 것이 아니라, 45개의 엔트리만을 고려한 최소값과 최대값으로 결정된다. 결국 (그림 6(a))의 엔빌로프 $[L^{[0,44]},U^{[0,44]}]$는 (그림 4(a))의 엔빌로프 $[L,U]$에 비해 좁게 결정되고, 이 엔빌로프에 의해 계산되는 하한 $LB_{[0,44]}(Q,S)$는 $LB_{SE}(Q,S)$에 비해 그 값이 커지게 된다. 실제 계산 결과를 비교하면, $LB_{[0,44]}(Q,S)$ = 81.7로 $LB_{SE}(Q,S)$ = 0에 비해 훨씬 크며, $RID(Q,S)$ = 137.8에 가까운 값을 가짐을 알 수 있다. 따라서, 이러한 $LB_{[a,b]}(Q,S)$을 매칭에 사용할 수 있다면, 전지 효과를 크게 발휘할 수 있게 된다. ■

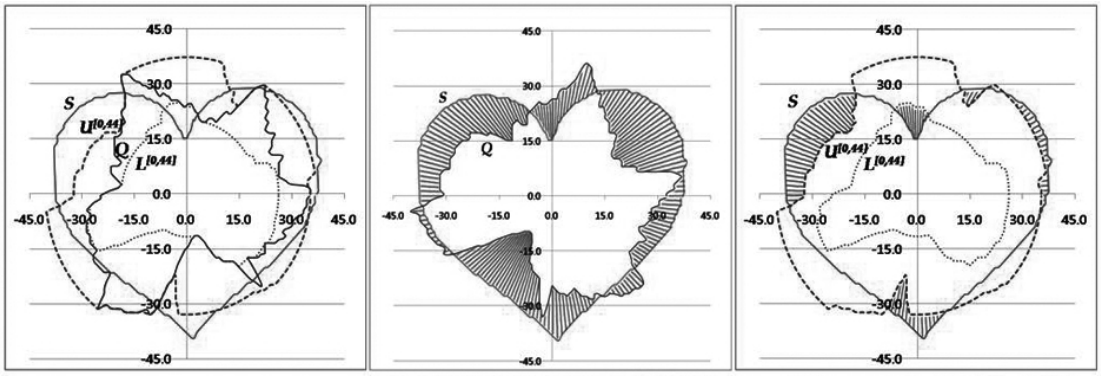
(그림 7)은 [예제 2](그림 6)의 시퀀스를 이미지 도메인에서 나타낸 것이다. (그림 5)에서와 마찬가지로, 나뭇잎 윤곽선은 질의 시퀀스 $Q$, 하트 윤곽선은 데이터 시퀀스이다. 그리고 나뭇잎 윤곽선을 포함하고 있는 점선은 엔빌로프 $U^{[0,44]}$를, 나뭇잎 윤곽선 안에 포함된 점선은 엔빌로프 $L^{[0,44]}$를 각각 나타낸다. (그림 7(b))의 빗금친 부분은 회전-불변 거리 $RID(Q^{[0,44]},S)$이며, (그림 7(c))에서 빗금친 부분은 이의 하한 $LB_{[0,44]}(Q,S)$를 나타낸다. (그림 5(c))에서 빗금친 부분이 없는 것과는 달리, (그림 7(c))에서는 하트 윤곽선의 데이터 이미지가 나뭇잎 윤곽선 이미지 엔빌로프 $[L^{[0,44]},U^{[0,44]}]$에 포함되지 않으며, 그 차이가 하한 $LB_{[0,44]}(Q,S)$로 계산(그림 7(c)에서 빗금친 부분)됨을 알 수 있다.

회전 구간이 [a,b]인 질의 시퀀스 $Q$의 하한 $LB_{[a,b]}(Q,S)$는 $RID(Q^{[a,b]},S)$의 하한이나 $RID(Q,S)$의 하한은 아닌 문제점이 있다. 즉, [예제 2]의 $LB_{[0,44]}(Q,S)$가 $LB_{SE}(Q,S)$보다 커서 전지 효과가 잘 나타나지만, 이 $LB_{[0,44]}(Q,S)$를 $RID(Q,S)$의 하한으로는 사용할 수는 없음을 뜻한다. 그 이유는 하한 $LB_{[0,44]}(Q,S)$가 $RID(Q^{[0,44]},S)$의 하한이기는 하나 $RID(Q,S)$의 하한은 아니기 때문이다. 예를 들어,




(a) 시퀀스와 엔빌로프. (b) 회전-불변 거리 $RID(Q^{[0,44]},S)$. (c) 엔빌로프 기반 하한 $LB_{[0,44]}(Q,S)$.

(그림 6) 회전 구간 [0,44]를 고려한 경우의 엔빌로프와 이의 하한




(a) 윤곽선 이미지와 엔빌로프. (b) 회전-불변 거리 $RID(Q^{[0,44]},S)$. (c) 엔빌로프 기반 하한 $LB_{[0,44]}(Q,S)$.

(그림 7) 이미지 도메인에서 본 회전 구간을 고려한 엔빌로프와 이의 하한

$RID(Q^{[0,44]},S)$가 $RID(Q^{[45,359]},S)$보다 클 수 있고, 이 경우 $RID(Q,S)$는 $RID(Q^{[0,44]},S)$가 아닌 $RID(Q^{[45,359]},S)$로 결정될 수 있기 때문이다. 이러한 문제를 해결하기 위해, 전체 회전 구간을 여러 개의 서로소(disjoint)인 회전 구간으로 나누고, 이들 회전 구간의 하한들의 최소값을 전체 구간의 하한으로 삼는 다중 엔빌로프 기법을 제안한다.

**[정리 1]** 길이 $n$인 질의 시퀀스 $Q$가 주어졌을 때, 전체 구간 [0,n-1]을 서로소인 m개의 회전 구간 $[a_0,b_0]$, $[a_1,b_1]$, …, $[a_{m-1},b_{m-1}]$ ($a_0=0$, $a_k=b_{k-1}+1$, $b_{m-1}=n-1$, $k=1,...,m-1$)로 나누었다 하자. 그러면, 각 회전 구간의 하한 $LB_{[a_k,b_k]}(Q,S)$의 최소값($=\min_{k=0}^{m-1} LB_{[a_k,b_k]}(Q,S)$)은 질의 시퀀스 $Q$와 데이터 시퀀스 $S$의 회전-불변 거리인 $RID(Q,S)$의 하한이다.

**[증 명]** 질의 시퀀스 $Q$의 모든 회전 시퀀스 중 데이터 시퀀스와의 거리가 최소인 회전 시퀀스를 $Q^j$라 하자. 그러면, $RID(Q,S)=RID(Q^j,S)$가 성립한다. 또한, 전체 구간을 서로소인 회전 구간들로 나누었기 때문에, $j$는 어느 한 회전 구간에 속하게 되는데, 이때 $j$가 속하는 회전구간을 [a,b]라 하자(즉, $a \le j \le b$). 그러면, $RID(Q,S)=D(Q^j,S)$이므로, 자연히 $RID(Q^{[a,b]},S)=D(Q^j,S)=RID(Q,S)$가 성립한다. 따라서, $LB_{[a,b]}(Q,S)$는 $RID(Q^{[a,b]},S)$의 하한인 동시에 $RID(Q,S)$의 하한이 된다. 결국, 이들 회전구간의 하한의 최소값인 $\min_{k=0}^{m-1} LB_{[a_k,b_k]}(Q,S)$는 $RID(Q,S)$의 하한이 된다. ■

본 논문에서는 [정리 1]의 하한 $\min_{k=0}^{m-1} LB_{[a_k,b_k]}(Q,S)$를 "여러 회전 구간에 대해 엔빌로프를 구성한다"는 의미에서 $LB_{ME}(Q,S)$로 표기한다. (ME는 multiple envelopes를 의미함.) [예제 2]에서 설명한 바와 같이, 일부 구간 [a,b]만을 고려한 하한 $LB_{[a,b]}(Q,S)$는 $LB_{SE}(Q,S)$보다 크게 되고, $LB_{[a,b]}(Q,S)$들의 최소값으로 계산된 $LB_{ME}(Q,S)$ 역시 $LB_{SE}(Q,S)$보다 커질 가능성이 높다. 따라서, 회전 불변 윤곽선 이미지 매칭에서 $LB_{SE}(Q,S)$ 대신 $LB_{ME}(Q,S)$를 사용하면, 불필요한 회전-불변 거리 계산을 더욱 줄일 수 있고, 이에 따라 성능을 크게 향상시킬 수 있다. (그림 8)은 다중 엔빌로프를 사용한 회전-불변 윤곽선 이미지 매칭 알고리즘 *RI-ME*를 나타낸다.

Algorithm *RI-ME*(query sequence $Q$, a set $S$ of data sequences, tolerance $\varepsilon$)
1. Divide the whole range [0,*n*-1] to *m* rotation ranges $[a_0,b_0]$, …, $[a_{m-1},b_{m-1}]$ ;
2. Construct $L^{[a_k,b_k]}$ and $U^{[a_k,b_k]}$ from $Q$ for each $[a_k,b_k]$ $(k=0,...,m-1)$ ;
3. $R := \emptyset$ ;
4. **for each** data sequence $S \in S$ **do**
5. **if** $LB_{ME}(Q,S) \le \varepsilon$ **then** // $LB_{ME}(Q,S)=\min_{k=0}^{m-1} LB_{[a_k,b_k]}(Q,S)$
6. **if** $RID(Q,S) \le \varepsilon$ **then**
7. $R := R \cup \{S\}$ ;
8. **end-if**
9. **end-if**
10. **end-for**
11. **return** R ;

(그림 8) 다중 엔빌로프 기반의 하한을 이용한 회전-불변 이미지 매칭 알고리즘

## 5. 다중 엔빌로프 기법에서 회전 구간 분할

다중 엔빌로프 기법을 사용하기 위해서는 전체 구간을 여러 회전 구간으로 나누는 방법이 필요하다. 본 절에서는 회전 구간 개수가 주어졌을 때, 각 회전 구간의 크기를 결정하는 방법을 논의한다. 제5.1절에서는 가장 간단하게 전체 구간을 동일한 너비로 나누는 동일 너비 분할을 제시한다. 다음으로, 제5.2절에서는 엔빌로프의 넓이를 최소화하는 방향으로 회전 구간을 구성하는 엔빌로프 최소화 분할을 제시한다. 마지막으로, 제5.3절에서는 엔빌로프 최소화 분할을 실용적으로 사용할 수 있는 휴리스틱 알고리즘을 제시한다.

### 5.1 동일-너비 분할

동일-너비 분할(equi-width division)은 전체 구간을 동일한 크기의 회전 구간으로 나누는 방법이다. 시퀀스의 길이가 $n$ 이고 회전 구간의 수가 $m$ 이라 할 때, 이 분할 기법은 전체 구간 $[0,n-1]$ 을 크기가 $n/m$ 인 $m$ 개의 회전 구간으로 나눈다. 좀 더 정확히 말해서, 회전 구간 $[0,n-1]$ 을 다음 식 (7)과 같이 $m$ 개 회전 구간으로 나눈다.

$$\left[0,\left\lfloor\frac{n}{m}\right\rfloor-1\right],\left[\left\lfloor\frac{n}{m}\right\rfloor,2\cdot\left\lfloor\frac{n}{m}\right\rfloor-1\right],\ldots,\left[(m-1)\cdot\left\lfloor\frac{n}{m}\right\rfloor,n-1\right] \quad (7)$$

동일-너비 분할은 매우 간단하다는 장점이 있다. 그러나, 하한 $LB_{ME}(Q,S)$ 를 증가시키고자 하는 특별한 노력이 사용된 바 없으며, 특히 시퀀스의 최대 및 최소값이 여러 구간에 산재하는 경우 하한이 작아지는 단점이 있다.

### 5.2 엔빌로프 최소화 분할

본 절에서는 엔빌로프의 넓이가 작아지면 하한이 커질 것이라는 점에 착안하여 엔빌로프 최소화(envelope minimization) 분할을 제안한다. 다중 엔빌로프 분할에 있어서, 엔빌로프 최소화 분할은 회전 구간들이 생성하는 엔빌로프들의 면적 합을 최소화하고자 하는 분할 기법이다. 이를 설명하기 위해, 먼저 엔빌로프 넓이를 다음과 같이 정의한다.

**[정의 6]** 길이가 $n$ 인 질의 시퀀스 $Q$ 가 주어졌을 때, 회전 구간 [a,b]의 엔빌로프 넓이 $Area(Q,[a,b])$ 는 $Q$ 의 엔빌로프 최소 및 최대 시퀀스 $L^{[a,b]}$ 과 $U^{[a,b]}$ 사이의 유클리디안 거리인 $\sqrt{\sum_{i=0}^{n-1}\left|l_i^{[a,b]}-u_i^{[a,b]}\right|^2}$ 로 정의한다. ■

[정의 6]의 엔빌로프 넓이를 사용하여, 엔빌로프 최소화 분할을 정형적으로 정의하면 다음과 같다.

**[정의 7]** 전체 회전 구간 $[0,n-1]$ 을 서로소인 $m$ 개의 회전 구간 $[a_0,a_1-1]$, $[a_1,a_2-1]$, …, $[a_{m-1},a_m-1]$ 으로 구성한다 하자( $a_0=0$, $a_m=n$ ). 이때, 회전 구간 $[a_k,a_{k+1}-1]$ 들의 엔빌로프 넓이의 합 $\sum_{k=0}^{m-1}Area(Q,[a_k,a_{k+1}-1])$ 이 최소화되도록 회전 구간들을 결정하는 방법을 엔빌로프 최소화 분할이라 정의한다. ■

그런데, 엔빌로프 최소화 분할은 회전 구간들을 구성하는 경우의 수가 너무 많아, 실용적으로 사용하기 어려운 문제점이 있다. 좀 더 자세히 설명하면, 전체 구간이 $[0,n-1]$ 인 경우, $m$ 개의 회전 구간을 구성하는 방법은 총 $\binom{n-2}{m-1}$ 로서 (0과 $n-1$ 을 제외한 $n-2$ 개의 수에서 총 $m-1$ 개의 분할 기준을 선택하는 경우의 수에 해당함), 그 경우의 수가 너무 많아지게 된다. 예를 들어, 길이( $n$ )가 360이고, 이를 8개의 회전 구간으로 분할한다 하면, 총 $\binom{358}{7}$ = 약 $1.41\times10^{14}$ 개의 많은 경우의 수가 발생하고, 이들 경우의 수 각각에 대해서 엔빌로프 넓이의 합을 구해야 하므로, 실용적으로는 사용할 수 없는 방법이다.

### 5.3 엔빌로프 최소화 분할의 휴리스틱 알고리즘

본 절에서는 엔빌로프 최소화 분할에서 회전 구간을 찾는 경우의 수를 줄이기 위한 휴리스틱 알고리즘을 제안한다. 제안하는 휴리스틱 알고리즘은 $\sum_{k=0}^{m-1}Area(Q,[a_k,a_{k+1}-1])$ 을 최소화하는 최적의 해답 대신에, 국부적(local) 최적 값을 찾는 방법으로 가능한 경우의 수를 크게 줄이면서 근사적 해답을 찾는 방법이다. (그림 9)는 엔빌로프 최소화 분할의 근사적 해답을 찾는 휴리스틱 알고리즘이다. 알고리즘을 설명하면 다음과 같다. 먼저 Line 1은 초기 분할을 수행하는 것으로, 우선 전체 회전 구간 [0,n-1]을 $m$ 개의 동일 너비 회전 구간으로 나눈다. Line 4~6에서는 각 구간의 경계인 $a_j$ 를 제한된 범위인 $a_{j-1}+1$ ( $a_{j-1}$ 보다는 커야 함)과 $a_{j+1}-2$ ( $a_{j+1}-1$ 보다는 작아야 함) 사이에서 이동시키면서 엔빌로프 넓이의 합을 구하되, 그 합을 최소로 하는 $a_j$ 를 찾아 새로운 경계로 삼는다. 다시 말해서, 이는 $a_j$ 가 가질 수 있는 범위 값인 $a_{j-1}+1$ 과 $a_{j+1}-2$ 사이에서 $a_j$ 의 국부적 최적

값을 찾는 과정이다. 이러한 과정은 $a_0$을 제외한 모든 $a_j$에 대해서 반복 수행하며, Line 4~6의 이러한 과정은 모든 $a_j$에 변화가 없을 때까지 반복 수행된다(Line 3~7).

Algorithm *Divide-Ranges*(query sequence $Q$)
1. Divide $[0,n-1]$ into $m$ partitions $[a_0,a_1-1]$, ..., $[a_{m-1},a_m-1]$ by the *equi-width division*:
2. Compute $\sum_{k=0}^{m-1} Area(Q,[a_k,a_{k+1}-1])$;
3. **repeat**
4. **for** $j:=1$ to $m-1$ **do**
5. Set $a_j$ as the value in the range of $a_{j-1}+1$ to $a_{j+1}-2$ such that $\sum_{k=0}^{m-1} Area(Q,[a_k,a_{k+1}-1])$ becomes minimum;
6. **end-for**
7. **until** no change;
8. **return** $[a_0,a_1-1]$, ..., $[a_{m-1},a_m-1]$;

(그림 9) 엔빌로프 최소화 기법의 휴리스틱 알고리즘

엔빌로프 최소화 분할의 휴리스틱 알고리즘을 사용하면 회전 구간을 구성하는 경우의 수를 크게 줄일 수 있다. (그림 9)의 휴리스틱 알고리즘의 계산 복잡도는 다음과 같다. Line 5를 보면, $a_j$에 대해서 $a_{j-1}+1$과 $a_{j+1}-2$사이의 모든 값을 대입하여 각 구간의 엔빌로프 넓이를 구한다. 여기서, $a_j$값들이 취할 수 있는 값은 결과적으로 1과 n-2 사이의 값들로서 총 n-2개가 되고, 이에 따라 Line 5의 계산 복잡도는 $\Theta(n)$이 된다. 그런데, Line 4~6의 for 루프는 총 m-1번 수행되므로, Line 4~6의 계산 복잡도는 $\Theta(mn)$이 된다. 만일, Line 3~7의 repeat 루프가 r번 수행된다고 하면, 휴리스틱 알고리즘의 전체 계산 복잡도는 $\Theta(rmn)$이 된다. 실험 결과, r은 n에 비해서 훨씬 작은 값(대략 10회 미만)으로 나타났으며, 따라서 $\Theta(rmn)$는 앞서의 $\Theta\left(\binom{n-2}{m-1}\right)$에 비해 계산 복잡도를 크게 줄인 것이라 할 수 있다. 예를 들어, n이 360이고, m이 8이라면, 앞서 최적 회전 구간 분할을 찾기 위해서는 $1.41\times10^{14}$개의 경우의 수를 고려해야 했으나, 휴리스틱 알고리즘에서는 대략 20,000개~30,000개의 경우의 수를 고려하는 것으로 나타났다.

## 6. 성능 평가

### 6.1 실험 환경 및 데이터

실험에서는 세 가지 데이터 집합을 사용하였다. 첫 번째 데이터 집합은 기존 연구[3,9]에서 자주 사용된 Mixed-bag으로, 총 160개의 이미지로 구성되어 있다. 이 데이터 집합을 *MIXED_DATA*라 한다. 두 번째 데이터 집합은 해양 생물 이미지로 구성된 SQUID 데이터 집합[21]으로서, 총 1,100개의 이미지로 구성되어 있다. 이 데이터 집합은 이미지 유사 검색에 사용되도록 공개된 것으로서, 본 논문에서는 이를 *SQUID_DATA*라 부른다. 세 번째 데이터 집합은 웹 상에서 직접 구한 이미지들로서 총 10,259개의 이미지로 구성되어 있다[7]. 본 논문에서는 이 데이터 집합을 *WEB_DATA*라 부른다. 실험에서는 우선 각 이미지에서 윤곽선을 추출한 후, 이를 길이 360의 시계열로 변환하여 데이터베이스를 구축하였다.

실험을 수행한 하드웨어 플랫폼은 UltraSPARC IIIi CPU 1.34GHZ, 1.0GB RAM, 80GB 하드 디스크를 장착한 SUN Ultra이며, 소프트웨어 플랫폼은 Solaris 10 운영 체제이다. 실험은 기존 방법인 *RI-Naïve*와 *RI-EA*, 그리고 본 논문에서 제안한 *RI-SE*와 *RI-ME*의 네 가지 알고리즘을 대상으로 하였다. 특히, 다중 엔빌로프 기반 매칭 알고리즘인 RI-ME에 대해서는, 동일-너비 분할을 사용하는 경우를 *RI-ME(EW)*, 엔빌로프 최소화 분할을 사용하는 경우를 *RI_ME(EM)*으로 표기한다(*EW*는 equi-width를, *EM*은 envelope minimization을 각각 나타낸다). 결국, 총 다섯 가지 알고리즘을 고려하였으며, 크게 네 가지 실험을 수행하였다. 첫 번째 실험은 RI-ME에서 회전 구간 수를 결정하는 실험이다. RI-ME의 경우 회전 구간 수에 따라 매칭 성능이 달라질 수 있다. 따라서 이 실험에서는 회전 구간 수가 얼마일 때 RI-ME가 가장 좋은 매칭 성능을 보이는지를 실험한다. 두 번째에서 네 번째 실험에서는 세 개의 데이터 집합 각각에 대해, 회전-불변 거리 계산 횟수와 실제 수행 시간을 측정한다. 회전-불변 거리 계산 횟수는 제안한 알고리즘이 불필요한 회전-불변 거리 계산을 얼마나 전지하는지를 알아보는 척도라 할 수 있다. 또한, 실제 수행 시간은 이러한 전지를 통해 얼마나 성능을 향상시켰는지를 나타내는 척도가 된다. 각 실험에서는 임의의 질의 시퀀스 10개를 선택하여, 이들 결과의 평균을 측정값으로 사용하였다.

### 6.2 실험 결과

#### 6.2.1 RI-ME의 회전 구간 개수 결정

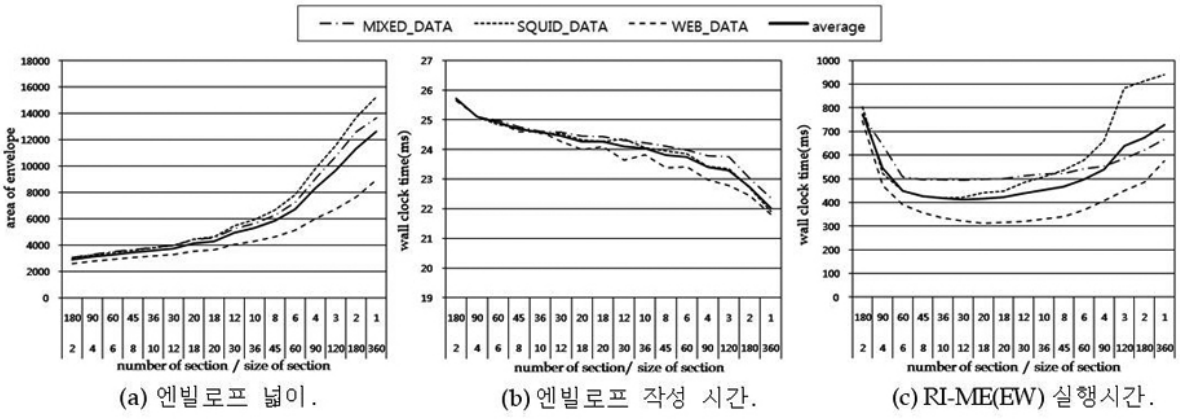
본 실험은 알고리즘 RI-ME에서 회전 구간 개수(즉, 회전 구간 크기)를 결정하기 위한 실험이다. RI-ME의 경우 회전 구간 수에 따라 성능이 달라질 수 있기 때문에, 다른 알고리즘과의 성능 비교에 앞서 최적의 회전 구간 수를 결정한다. 본 논문에서는 실험을 통해 RI-ME(EW)의 최적 회전 구간 개수를 구하고, 이를 RI-ME(EW)와 RI-ME(EM)의 회전 구간 수로 사용한다.

(그림 10)은 RI-ME(EW)에서 회전 구간의 개수(크기)에 따른 엔빌로프 넓이 변화, 하한 계산 시간, RI-ME(EW)의 실제 실행 시간을 측정한 결과이다. 실험에서 회전 구간 수는 180, 90, 60, …, 3, 2, 1로 달리 하였으며, 허용치는 60으로 고정하였다. (그림 10(a))를 보면 회전 구간 수가 줄어듦

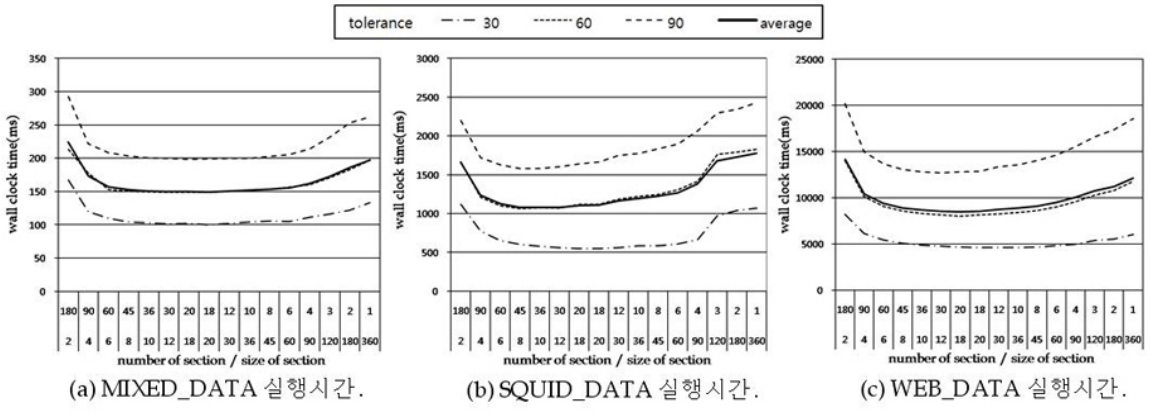
(회전 구간 크기가 넓어짐)에 따라 엔빌로프 넓이가 증가함을 알 수 있다. 이는 당연한 결과로, 회전 구간 수가 적을수록 한 구간에서 많은 수의 회전 시퀀스를 고려해야 하고, 이에 따라 엔빌로프가 넓어지기 때문이다. 제4절에서 설명했듯이 엔빌로프가 넓을수록 전지 효과가 떨어지므로, (그림 10(a))의 결과만 봤을 때 회전 구간 수가 많을수록 좋은 성능을 얻을 수 있게 된다. 그러나, 회전 구간 수를 많게 하면, 더 많은 수의 하한($\min_{k=0}^{m-1} LB_{[a_k,b_k]}(Q,S)$)을 계산해야 하는 오버헤드가 있다. (그림 10(b))가 이러한 하한 계산 오버헤드를 실행 시간으로 측정한 것이다. (그림 10(b))를 보면 회전 구간 수 m이 줄어들수록(회전 구간 크기가 넓어질수록) 하한 계산 시간이 줄어듦을 알 수 있다. 결국, (그림 10(a))와 (그림 10(b))의 결과를 종합해 보면, 회전 구간 개수가 적을수록 엔빌로프는 넓어지나 (하한은 증가하나) 하한 계산 시간은 줄어들고, 반대로 회전 구간 개수가 많을수록 엔빌로프는 좁아지나(하한은 감소하나) 하한 계산 시간은 늘어나는 상충 관계가 있음을 알 수 있다.

엔빌로프 넓이와 하한 계산 시간 사이의 상충관계에 따라, 본 논문에서는 RI-ME(EW)의 실제 매칭 시간을 바탕으로 최적의 회전 구간 수를 결정하고자 한다. 실제 실행 시간은 이들 두 팩터를 모두 반영한 결과로 볼 수 있기 때문이다. (그림 10(c))를 보면, (그림 10(a))와 (그림 10(b))를 합쳐 놓은 듯한 결과를 보이는데, 이는 실제 실행 시간이 엔빌로프 넓이와 하한 계산 시간에 모두 영향을 받기 때문이다. (그림 10(c))의 그래프가 전체적으로 U자 형태를 띄는 이유는 회전 구간 수가 많은 곳(그래프의 왼편)은 하한 계산에 많은 시간이 걸리고, 회전 구간 수가 적은 곳(그래프의 오른편)은 하한에 의한 미리 버림 효과가 작기 때문이다. (그림 10(c))의 실험 결과를 보면, 데이터 종류에 관계없이 회전 구간 수가 36, 30, 20, 18일 때 실제 실행 시간이 가장 짧은 것으로 나타났다.

(그림 11)은 허용치에 따른 RI-ME(EW)의 실제 실행 시간을 측정한 결과이다. 회전 구간 수는 (그림 10)과 동일하




(그림 10) 회전 구간 크기 따른 엔빌로프 작성과 RI-ME(EW)의 실행 시간 비교

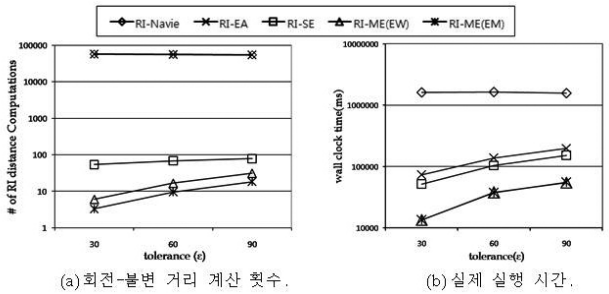



(그림 11) 허용치에 따른 RI-ME(EW)의 실행 시간 비교

게 하였으며, 허용치 30, 60, 90으로 하였다. (그림 11)을 보면, 허용치에 따라 그래프의 위치가 다르다. 이는 허용치가 커짐에 따라 많은 이미지가 유사하다고 판단되어 회전-불변 거리를 계산하는 횟수가 많아지기 때문이다. 주목할 점은 (그림 11)의 모든 그래프 모양이 (그림 10(c))와 비슷한 U자의 형태를 보인다는 점이다. 앞서 설명했듯이, 이는 RI-ME(EW)의 실제 매칭 시간이 엔빌로프 넓이와 하한 계산 시간 모두 영향을 받기 때문이다. 허용치를 달리했음에도 불구하고, 실행 시간이 가장 짧게 걸린 구간 수는 여전히 36, 30, 20, 18로 나타났다. 따라서, 본 논문에서는 이들 구간 수 중 첫 번째인 36을 RI-ME(EW)와 RI-ME(EM)의 회전 구간 수로 사용한다.

6.2.2 MIXED_DATA의 실험결과

(그림 12)는 MIXED_DATA에 대해 허용치를 달리하면서 회전-불변 거리 계산 횟수와 실제 실행 시간을 측정한 결과이다. (그림 12(a))를 보면, 제안한 RI-SE와 RI-ME(EW), RI-ME(EM)가 기존의 RI-Naïve와 RI-EA에 비해 회전-불변 거리 계산 횟수를 크게 줄였음을 알 수 있다. 이는 엔빌로프 기반의 하한 기법을 적용하는 본 논문의 접근법이 많은 불필요한 회전-불변 거리 계산을 전지할 수 있음을 의미한다. (그림 12(a))에서 RI-Naïve와 RI-EA가 동일하게 나타났는데, 이는 RI-EA의 경우 미리 버림을 통해 성능을 향상시키기는 하나(그림 12(b) 참조), 회전-불변 거리 계산 횟수 자체를 줄이지는 못하기 때문이다. 그리고, RI-ME(EW)와 RI-ME(EM)이 RI-SE에 비해 회전-불변 거리 계산 횟수를 더욱 줄였음을 알 수 있다. 이는 RI-ME에서 사용한 다중 엔빌로프 기반의 하한 계산이 단일 엔빌로프 기반의 하한 계산보다 우수한 전지 효과를 발휘함을 의미한다. 그리고 RI-ME(EW)보다 RI-ME(EM)이 더 효율적인 것을 볼 수 있다. 이는 엔빌로프 최소화 분할이 동일-너비 분할에 비해 엔빌로프 넓이를 줄이고 이에 따라 하한을 증가시키며, 궁극적으로 더욱 많은 전지 효과를 발휘함을 의미한다.



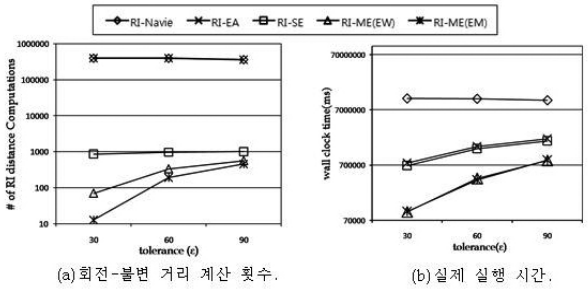

(a) 회전-불변 거리 계산 횟수. (b) 실제 실행 시간.

(그림 12) MIXED-DATA에서의 회전-불변 거리 계산 횟수와 실제 실행 시간

(그림 12(b))의 실행 시간 결과를 보면, 제안한 알고리즘이 기존의 알고리즘에 비해 성능을 크게 향상시킨 것으로 나타났다. (세로축이 log 스케일임에 유의한다.) 먼저 제안한 RI-SE와 RI-ME가 RI-Naïve 및 RI-EA에 비해 성능을 크게 향상시켰는데, 그 이유는 앞서 (그림 12(a))에서 보듯이 회전-불변 거리 계산 횟수를 크게 줄였기 때문이다. (그림 12(a))와는 달리, RI-EA가 RI-Naïve에 비해 우수한 성능을 보이는데, 이는 회전-불변 거리 계산에 있어 미리 버림[9]이 효과를 발휘하기 때문이다. 그러나, 모든 경우에 있어 제안한 RI-SE 및 RI-ME가 기존의 RI-Naïve는 물론 RI-EA에 비해 우수한 성능을 보이고 있다. 특히, 전지 효과가 가장 뛰어난 RI-ME(EM)이 가장 우수한 성능을 보였으며, 실제로 RI-ME(EM)은 RI-Naïve에 비해서는 119.2배, RI-EA에 비해서는 5.4배까지 성능을 향상시킨 것으로 나타났다.

6.2.3 SQUID_DATA의 실험결과

(그림 13)은 SQUID_DATA에 대한 실험 결과이다. (그림 13)의 실험 결과를 보면, 전반적인 경향이 (그림 12)와 매우 유사함을 볼 수 있다. 즉, 회전-불변 거리 계산의 횟수 및 수행 시간 모두에 있어서, 제안한 알고리즘이 기존의 알고리즘에 비해 좋은 결과를 보이고 있다. 그러나, (그림 12(a))와 (그림 13(a))를 비교해 보면, 허용치가 증가함에 따라 회전-불변 거리 계산 횟수의 차이가 빠르게 감소함을 알 수 있다. 이는 SQUID_DATA의 이미지 모양이 다양하지 못하여 허용치가 커짐에 따라 유사하다고 판단되는 이미지가 많아지기 때문이다. 또한, (그림 13(b))를 보면 RI-EA와 RI-SE의 수행 시간이 비슷한 결과를 보이는데, 이는 RI-SE의 문제점인 이미지 시퀀스의 변화폭이 크면 엔빌로프가 크게 구성되는 현상이 나타나기 때문이다. 즉, SQUID_DATA의 경우 이미지 시퀀스 변화 폭이 커서 RI-SE의 엔빌로프가 크게 구성되고, 결국 전지 효과가 줄어 들었기 때문이다. (그림 13)을 보면, 여전히 RI-ME가 RI-Naïve 및 RI-EA에 비해 우수하며, 특히 RI-ME(EM)은 RI-Naïve와 RI-EA에 비해 111.9, 7.5배까지 성능을 향상시킨 것으로 나타났다.



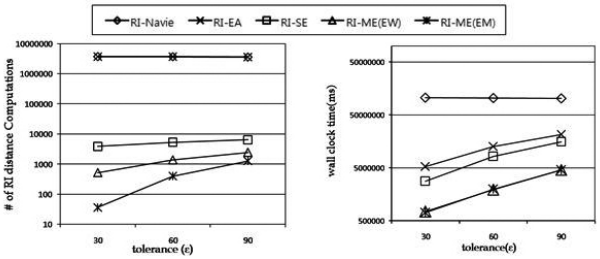

(a) 회전-불변 거리 계산 횟수. (b) 실제 실행 시간.

(그림 13) SQUID_DATA에서의 회전-불변 거리 계산 횟수와 실제 실행 시간

6.2.4 WEB_DATA의 실험결과

(그림 14)는 WEB_DATA에 대한 실험 결과를 나타낸다. (그림 14)의 결과를 보면, 전반적인 경향은 (그림 12) 및 (그

림 13)과 매우 유사함을 알 수 있다. 앞서의 실험과 마찬가지로, RI-Naïve, RI-EA, RI-SE, RI-ME 순의 성능을 보였으며, RI-ME(EM)이 가장 좋은 결과를 나타내었다. RI-ME(EM)은 RI-Naïve 및 RI-EA에 비해 최대 147.7배, 7.4배까지 성능을 향상시킨 것으로 나타났다. (그림 12~14)의 실험 결과를 종합하여, 데이터의 종류에 관계 없이 제안한 엔빌로프 기반 전지 기법이 기존 알고리즘에 비해 우수하다 할 수 있다.



(a) 회전-불변 거리 계산 횟수. (b) 실제 실행 시간.

(그림 14) WEB_DATA에서의 회전-불변 거리 계산 횟수와 실제 실행 시간

## 7. 결 론

본 논문에서는 회전-불변 윤곽선 이미지 매칭에 있어서, 불필요한 회전-불변 거리 계산을 크게 줄이는 효율적인 방법을 제안하였다. 제안한 방법은 엔빌로프 기반 하한을 사용하여 불필요한 계산을 줄이는 기법으로, 실험을 통하여 기존 회전 불변 이미지 매칭보다 훨씬 우수한 성능을 보이는 것을 확인하였다.

먼저, 단일 엔빌로프 기반 하한 기법을 제안한다. 단일 엔빌로프 기반 하한은 모든 회전 시퀀스를 고려하여 질의 시퀀스의 최고값을 엔트리로 가지는 시퀀스와 최소값을 엔트리로 가지는 시퀀스를 구한다. 이 두 개의 시퀀스를 엔빌로프라 정의하고, 엔빌로프와 데이터 시퀀스 간의 거리를 계산하여 불필요한 회전-불변 거리 계산을 전지해주는 하한이 됨을 [보조 정리 2]에서 증명한다.

다음으로, 다중 엔빌로프 기반의 하한 기법을 제안한다. 다중 엔빌로프 기반 하한은 단일 엔빌로프 기반 하한의 문제점인 질의 시퀀스의 변화폭이 큰 경우 하한이 줄어드는 문제점을 해결하기 위해 회전 구간이라는 개념을 도입하여 다중 엔빌로프 기반 하한으로 확장한다. 다중 엔빌로프는 질의 시퀀스 전체를 고려하지 않고, 서로소인 회전 구간만 고려하여 엔빌로프를 구성함으로써 하한이 줄어드는 문제를 해결한다. 또한 다중 엔빌로프 기반 하한에서 회전 구간 결정 방법으로 동일 너비 구간 방법과 엔빌로프 최소화 기법을 제안함으로 더욱 하한의 효과를 좋게 하였다.

제안하는 알고리즘의 성능을 평가하기 위해 실제 이미지 데이터 집합에 대한 실험을 통해 효과를 측정하였다. 실험 결과, 제안한 방법이 기존의 방법에 비해 최소 7배에서 최대 140배 좋은 효과를 보이는 것을 확인하였으며, 또한 단일 엔빌로프 기반 하한의 문제점을 다중 엔빌로프 기반 하한 방법이 해결하였음을 보여주었다.

## 참 고 문 헌

[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., “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.

[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] Kim, B.-S., Moon, Y.-S., and Kim, J., “Noise Control Boundary Image Matching Using Time-Series Moving Average Transform,” In Proc. of the 19th Int’l Conf. on Database and Expert Systems Applications (DEXA 2008), Turin, Italy, Sept. 2008

[8] 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.

[9] Keogh, E. J., Wei, L., Xi, X., Vlachos, M., Lee, S.-H., and Protopapas, P., “Supporting Exact Indexing of Arbitrarily Rotated Shapes and Periodic Time Series under Euclidean and Warping Distance Measures,” The VLDB Journal, Vol. 18, No. 3, pp. 611-630, June 2009.

[10] 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.
[11] Chan, K.-P., Fu, A. W.-C., and Yu, C. T., "Haar Wavelets for Efficient Similarity Search of Time-Series: With and Without Time Warping," IEEE Trans. on Knowledge and Data Engineering, Vol. 15, No. 3, pp. 686-705, Jan./Feb. 2003.
[12] Gonzalez, R. C. and Woods, R. E., Digital Image Processing, 2nd Ed., Prentice Hall, New Jersey, 2002.
[13] Pratt, W. K., Digital Image Processing, 4th Ed., Eastman Kodak Company, Rochester, New York, 2007.
[14] 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.
[15] 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.
[16] Wang, Z., Chi, Z., Feng, D., and Wang, Q., "Leaf Image Retrieval with Shape Features," In Proc. 4th Int'l Conf. on Advances in Visual Information Systems, Lyon, France, pp. 477-487, Nov. 2000.
[17] Zhang, D. Z. and Lu, G., "Review of Shape Representation and Description Techniques," Pattern Recognition, Vol. 37, No. 1, pp. 1-19, July 2003.
[18] Chen, y.-W. and Lee, S. C.,"The C-Chain Code - A New Method For Coding 3D Curves," In Proc. The 25th Asilomar Conf. on Signals, Systems and Computers, Vol.1, pp.472-476, Nov.1991.
[19] Grosky, W. I., and Mehrotra, R., "Index-Based Object Recognition in Pictorial Data Management," Computer Vision, Graphics, and Image Processing, Vol. 52, Issue3, pp.416-436, Dec.1990.
[20] Bebis, G., Papadourakis, G., Orphanoudakis, S., "Curvature Scale space driven Object Recognition with an Indexing Scheme Based on Artificial Neural Networks," Pattern Recognition, Vol.32, No.7, pp.1175-1201, July 1999.
[21] SQUID: http://www.ee.surrey.ac.uk/CVSSP/demos/css/demo.html

**김 상 필**

e-mail : spkim@kangwon.ac.kr
2009년 강원대학교 전산학과(학사)
2009년～현 재 강원대학교 컴퓨터과학과 석사과정
관심분야 : Data Mining & Knowledge Discovery, Data Mining Applications, Mutimedia Databases, Privacy Preserving Data Mining

**문 양 세**

e-mail : ysmoon@kangwon.ac.kr
1991년 한국과학기술원 전산학과(학사)
1993년 한국과학기술원 전산학과(석사)
2001년 한국과학기술원 전자전산학과 (박사)
1993년～1997년 현대전자산업(주) 주임연구원
2001년～2002년 ㈜현대시스콤 선임연구원
2002년～2005년 ㈜인프라밸리 기술위원(이사)
2005년～2008년 한국과학기술원 첨단정보기술연구센터 연구원
2008년～2009년 미국 퍼듀대학교 방문연구원
2005년～현 재 강원대학교 컴퓨터과학과 부교수
관심분야 : Data Mining, Knowledge Discovery, Stream Data, Storage System, Database Applications, Mobile/ Wireless Communication Services & Systems

**홍 선 경**

e-mail : hongssam@kangwon.ac.kr
1994년 강원대학교 전산학과(학사)
2004년 강원대학교 교육대학원 컴퓨터교육전공(석사)
2010년～현 재 강원대학교 컴퓨터과학 박사과정
1994년～1999년 한국정보문화센터 정보화교육지원본부 근무
2001년～2006년 한국정보통신대학원대학교 부설 정보통신교육원 (춘천분원) 강사
관심분야 : Data Mining & Knowledge Discovery, Privacy Preserving Data Mining, Computer Education