

# 중요도 기반 퍼지 원 클래스 서포트 벡터 머신을 이용한 비디오 요약 기술<sup>☆</sup>

## Video Summarization Using Importance-based Fuzzy One-Class Support Vector Machine

김기주\*                      최영식\*\*  
KiJoo Kim                      YoungSik Choi

### 요 약

본 논문에서는 비디오 요약을 시각적으로 특징이 있고 주관적으로 중요한 비디오 세그먼트 집합을 구하는 새로운 요약 방식을 기술한다. 시각적으로 특징이 있는 데이터 포인트를 찾기 위해 novelty detection으로 잘 알려져 있는 OC-SVM (One-Class Support Vector Machine)을 사용할 수 있다. 그러나 OC-SVM의 처리과정에 비디오 세그먼트에 대한 사용자의 주관적인 중요도를 반영하기는 어렵다. OC-SVM의 처리과정에 사용자의 주관적 중요성을 반영하기 위해서, 본 논문에서는 OC-SVM의 퍼지 버전을 유도한다. IFOC-SVM(Importance-based Fuzzy One-Class Support Vector Machine)은 비디오 세그먼트의 중요도에 따라 각 데이터 포인트에 가중치를 부여하고 데이터 분포의 서포트를 측정한다. 이때, 구해진 서포트 벡터는 비디오 세그먼트의 중요도와 시각적 특징 관점에서 비디오의 내용을 축약하여 표현한다. 제안된 알고리즘의 성능을 증명하기 위하여 가상의 데이터들과 다양한 종류의 비디오들을 가지고 실험하였다. 실험 결과는 제안하는 방법의 성능이 다른 비디오 요약의 성능보다 우수함을 보여주었다.

### ABSTRACT

In this paper, we address a video summarization task as generating both visually salient and semantically important video segments. In order to find salient data points, one can use the OC-SVM (One-class Support Vector Machine), which is well known for novelty detection problems. It is, however, hard to incorporate into the OC-SVM process the importance measure of data points, which is crucial for video summarization. In order to integrate the importance of each point in the OC-SVM process, we propose a fuzzy version of OC-SVM. The Importance-based Fuzzy OC-SVM weights data points according to the importance measure of the video segments and then estimates the support of a distribution of the weighted feature vectors. The estimated support vectors form the descriptive segments that best delineate the underlying video content in terms of the importance and saliency of video segments. We demonstrate the performance of our algorithm on several synthesized data sets and different types of videos in order to show the efficacy of the proposed algorithm. Experimental results showed that our approach outperformed the well known traditional method.

☞ keyword : 중요도 기반 퍼지 원 클래스 서포트 벡터 머신, 비디오 요약, Importance-based Fuzzy One-class SVM, IFOC SVM, One-class SVM, SVM, video summarization

## 1. 서 론

비디오 정보가 늘어남에 따라, 비디오 정보에 대한 효과적인 액세스 기술이 매우 중요하게 되었다. 이러한 인식은 “비디오 내용분석 및 검색 (Video Content Analysis and Retrieval)”이라고 일컫는 새로운 학문분야를 태동 시켰으며, 현재 매우 활발하게 연구되고 있는 학문분야로 자리

\* 정 회 원 : 한국항공대학교 대학원 컴퓨터공학과 박사과정  
zgeniez@kau.ac.kr

\*\* 종신회원 : 한국항공대학교 항공전자 및 정보통신공학부  
교수 choimail@kau.ac.kr(교신저자)

[2011/03/04 투고 - 2011/03/07 심사(2011/05/06 2차 - 2011/07/15 3차) - 2011/07/20 심사완료]

☆ 이 논문은 2011년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 부분적으로 받아 수행된 기초연구사업(No. 2011-0012297).

잡게 되었다[1].

비디오 내용분석 및 검색에 있어서 비디오 초록 생성은 (Video Abstraction) 원본 비디오를 자동 축약하는 기술로서, 방대한 비디오 정보를 검색/브라우징 하는데 필수적인 기술로서 가장 활발히 연구가 수행되고 있는 분야 중의 하나이다[1]. 비디오 초록 생성 기술은, 크게 “비디오 훑어보기”(Video Skimming), “주요장면 추출하기”(Video Highlights), “비디오 요약하기”(Video Summarization)로 구분 될 수 있다.

본 논문에서 다루려는 “비디오 요약하기”는 비디오에 대한 구조적이고 의미적인 정보를 추출하여 찾는 것이다[2-5]. 비디오 요약의 결과는 일반적으로 주요정지영상의 시퀀스로(sequence of keyframe) 표현된다. 키프레임(key frame)으로 표현된 비디오 요약은 비디오 전체의 내용을 한눈에 파악할 수 있게 할 뿐만 아니라, 그 키프레임을 담고 있는 장면의 창구(entrances) 역할을 하게 된다. 이 때문에, 비디오 요약을 최적의 키프레임 선정, 혹은 최적의 키프레임이 위치한 세그먼트(segment) 선정 작업이라고 볼 수 있다.

본 논문에서는 비디오 요약을 “시각적으로 특징이 있고, 주관적으로 중요한 최적의 세그먼트들을 선정하는 것”으로 새롭게 해석하였다. 즉, 비디오 시퀀스를 “시각적인 특징 요소인 칼라, 모션, 또는 시간에 따른 변화 등과 같은 관점에서 특징이 있고”, “사람의 주관적인 판단에 의해 중요한” 세그먼트들의 최적의 집합으로써 간주하였다. 시각적으로 특징 있는 세그먼트들을 얻는 문제는 “novelty detection” 문제로 해석할 수 있으며, OC-SVM(One Class Support Vector Machine) 방법을 통해 해결할 수 있다. OC-SVM은 데이터 전체에 대한 분포를 추정하는 것이 아니라, 데이터 분포의 서포트를 통계적으로 추정하는 기술로서 주어진 데이터를 가장 잘 표현할 수 있는 최적의 서포트 벡터를 구하는 기술이다. 이러한 OC-SVM은 “novelty detection” 문제와 이미지 검색에 대하여 성공적으로 사용 되어 졌다[6-8].

OC-SVM이 “novelty detection”에 대하여 잘 알려진 방법이지만, OC-SVM에 사람의 중요도를 결합하는 것은 어려운 일이다. 왜냐하면, OC-SVM은 모든 데이터들을 동등하게 다루기 때문이다. 본 논문에서 새롭게 해석한 비디오 요약에서는 시각적 특징뿐 만 아니라, 사용자의 주관적인 중요도 역시 매우 중요하다. 예를 들면, “길이가 긴 세그먼트가 짧은 것 보다 중요하다.”, “사람의 얼굴이 담겨 있는 세그먼트는 매우 중요하다.” 등과 같이 사용자의 주관적인 중요도가 비디오 요약에 반영 되어져야 하기 때문이다. 따라서 OC-SVM에 세그먼트의 주관적인 중요도를 결합하기 위한 방법이 필요하다.

[9-11]에서는 OC-SVM의 퍼지 버전을 유도하였다. 그러나 [9]에서는 데이터들이 서포트 벡터가 될 때에만 퍼지 멤버십이 적용된다. 즉, 에러를 발생하지 못한 데이터는 중요도가 높다 하여도 서포트 벡터로 선정 될 가능성이 전혀 없다. 반대로 말하면 서포트 벡터들의 멤버십만이 적용된 것이다. [10]에서는 데이터들의 시각적 특징과 퍼지 멤버십을 같이 다루고 있지만, 서포트 벡터로 선정되는 것이 시각적인 특징이나 퍼지 멤버십 하나만으로 결정될 수도 있다. [11]에서는 데이터들의 벡터 크기를 퍼지 멤버십에 따라 조절 한 후, 초월구(hyper sphere)를 이용한 OC-SVM에 적용하였지만, 데이터들의 퍼지 멤버십과 분포에 따라 초월구의 중심 벡터 크기가 반지름 보다 큰 경우와 작은 경우로 나누어지고, 이에 따라 퍼지 멤버십이 서포트 벡터 선정에 주는 영향이 달라진다.

이러한 퍼지 기반 OC-SVM들의 특징들은 본 논문에서 제안하는 비디오 요약에는 적합하지 않다. 따라서, 본 논문에서는 모든 데이터들의 시각적 특징과 퍼지 멤버십을 같이 고려하는 중요도 기반 퍼지 원 클래스 서포트 벡터 머신(Importance-based Fuzzy OC-SVM, 줄여서 IFOC-SVM)을 유도 한다.

IFOC-SVM의 특징은 모든 데이터의 중요도를 고려하여 데이터 분포의 서포트를 측정한다. 이

때, 구해진 서포트 벡터는 비디오 세그먼트의 중요도와 시각적 특징 관점에서 비디오의 내용을 축약하여 표현한다. 이러한 특징을 이용하여 IFOC-SVM을 반복적으로 수행함으로써 “scalable video summarization”의 구조를 이룰 수 있다. 제안된 알고리즘은 성능을 증명하기 위하여 다양한 비디오를 가지고 실험 하였다.

본 논문의 다음 부분은 다음과 같이 구성되어 있다. 2장에서는 비디오 요약과 OC-SVM의 관련 연구들을 소개하고, 3장에서는 OC-SVM의 이론과 성질에 대해 언급한다. 4장에서는 본논문에서 제안하는 IFOC-SVM을 유도하고 분석하고, 5장에서 IFOC-SVM을 이용한 비디오 요약방법에 대해서 설명한다. 6장에서는 가상 데이터들과 다양한 비디오 데이터들로 실험을 하고 그 결과를 분석한다. 마지막으로 7장에서는 결론 및 향후 연구 과제를 제시한다.

## 2. 관련 연구

### 2.1 비디오 요약 기술

비디오 요약 알고리즘은 상당히 많다. 이들 비디오 요약의 목적에 따라 크게 두 카테고리로 나눌 수 있다.

첫 번째 비디오 요약 알고리즘 카테고리에서는 하나의 샷(shot)에서 시각적인 움직임을 잘 표현하는 키프레임을 선정 한다[12-14]. 이러한 방법은 하나의 짧은 비디오에서 시각적인 내용의 비디오 요약에 대해서는 효과적일 수는 있어도, 긴 비디오에 대해서는 좋은 성능이 나오지 못한다. 왜냐하면 비디오의 내용을 빠르게 인식하기에는 키프레임 수가 너무 많기 때문이다. 이러한 문제점들을 극복하기 위해서, 최근 연구들은 긴 비디오에 대한 요약 알고리즘을 연구하기 시작하였다[2,3,5,15]. 이러한 접근의 기본적인 요약 방법은 우선 비디오를 기본적인 단위로 분할한다. 그런 후, 비디오 요약 알고리즘은 세그먼트의 성질을 기반으로 대표적인 세그먼트들을 선정한다.

다. 각각의 추출된 세그먼트들의 중심은 키프레임으로 선정되어지고, 추출된 키프레임들은 events, scene, 또는 stories등과 같이 비디오의 기본 단위로 표현한다.

[2]의 Yeung과 Yeo는 우선 비디오를 샷으로 분할하고, 샷을 그룹화 하는데 “time-constrained clustering”을 적용하여 모든 샷에 대해서 어떤 클러스터에 속해 있는지 표시하였다. 그들은 샷의 클러스터 넘버를 비디오 스토리의 기본 단위와 “대화”와 “동작”등의 사건을 식별하기 위해 사용하였다. 또한, 비디오 스토리의 기본 단위에 대하여 비디오 포스터(video poster)라 불리는 pictorial 비디오 요약을 개발하였다. 스토리의 기본 단위에 대한 클러스터 장악력은 클러스터에 존재하는 샷의 수 또는 클러스터에 존재하는 샷의 길이에 비례하였다. 이 장악력은 비디오 포스터의 이미지 크기를 결정했다. 이 방법은 스토리의 기본 단위 검색 방법 (혹은 신 추출 기술)에 의존적이다.

[3]의 Uchihashi등은 비디오 요약을 중요한 세그먼트를 선정하는 것으로 정의했다. 그들은 세그먼트의 길이와 희소성을 중요도로 사용했다. 이 방법에서, 모든 비디오 프레임은 미리 정의된 클러스터의 수에 의한 계층적 클러스터링을 사용하여 분할되어진다. 클러스터의 연속적인 프레임들은 하나의 세그먼트로 묶여진다. 세그먼트들은 중요도로써 순위가 붙여진다. 비디오 요약은 실험적으로 정해진 세그먼트의 수를 가지고 만들어진다. 키프레임은 각 세그먼트의 중심에 가장 가까운 프레임이 추출된다. 그들은 pictorial 비디오 요약을 하기 위한 frame-packing 알고리즘을 제안하였다. [5]의 Andreas Girgensohn등은 이 비디오 요약 방법의 가능한 응용을 보여주었다.

[15]의 Hanjalic과 Zhang은 비디오 요약을 클러스터의 최적 수를 찾는 것으로 해석하였다. 그들은 비디오를 2~N개의 클러스터로 분할하였다. 여기서, N은 비디오의 길이에 비례한다. 그들은 클러스터의 최적의 집합을 결정하는 cluster validity measure를 제안하였다. 그러나 클러스터의 수를 결정하는 것은 아직도 어려운 일이다.

## 2.2 OC-SVM 기술

많은 응용분야에서 주어진 데이터의 분포를 측정하는 것은 매우 어려운 문제이다. 이를 극복하기 위하여, OC-SVM은 데이터 분포의 서포트를 측정함으로써 문제를 단순화 한다[6,8]. 즉, OC-SVM은 데이터가 존재하는 영역을 추출하며, 이 영역을 서포트 벡터로써 표현한다. 표현된 영역 밖의 데이터 포인트들은 아웃라이어(outlier) 간주된다.

OC-SVM은 “클러스터링”(clustering)[16], “novelty detection”[8,17,18], “내용기반 이미지 검색(content-based image retrieval)”[7,19] 등과 같은 다양한 응용분야에 쓰이고 있다. [17]에서는 OC-SVM을 갑작스런 스펙트럼 변화 검출에 사용하였고, 기존의 알고리즘보다 우수한 방법이라는 것을 보였다. [8]에서는 미국 우편 서비스 데이터베이스로부터 필기체 숫자(handwritten digits)의 아웃라이어 검출에 OC-SVM을 사용하였다. [9]에서는 “사용자 의사반영 처리(relevance feedback process)”를 이용한 이미지 검색 시스템을 구축하였다.

또한, OC-SVM의 결과로 얻어진 서포트 벡터는 전체 데이터 포인트들의 분포를 나타내는 표현 벡터로(descriptive vector) 해석할 수 있다. 즉, OC-SVM은 대부분의 데이터 포인트들이 존재하는 영역의 표면에 있는 벡터들을 선택하고, 선택되어진 서포트 벡터들을 가지고 영역을 표현한다. OC-SVM의 이러한 성질은 이미지 검색에서 사용되었다[7,19]. 즉, 서포트 벡터들이 데이터의 분포의 윤곽을 묘사한다는 것이 OC-SVM의 주목할 만한 성질이다.

## 3. One-Class SVM

[8]에서는 모든 데이터 포인트가 존재하는 영역을 평면  $\omega \cdot \phi(x) - \rho = 0$ 로 구분한다. 이를 식으로 표현하면 다음과 같다.

$$\begin{aligned} & \text{minimize} \quad \frac{1}{2} \|\omega\|^2 + C \sum_i \xi_i - \rho & (1) \\ & \text{subject to} \quad \omega \cdot \phi(x) \geq \rho - \xi_i, \xi_i \geq 0, \forall i \end{aligned}$$

여기서,  $C$ 는 상수,  $\xi_i$ 는 패널티(penalty),  $\|\cdot\|$ 는 유클리디안 놈(Euclidean norm)이고  $\phi(x)$ 는  $x$ 를 고차원 특징 공간인 커널 공간(kernel space)로 사상(mapping)하는 사상함수 이다[8,15]. 식 (1)은 KKT 조건이 적용된 라그랑지안 계수법을 사용하여 풀 수 있고, 다음의 식을 얻을 수 있다.

$$\begin{aligned} & \text{maximize} \quad W = -\frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j K(x_i, x_j) & (2) \\ & \text{subject to} \quad \sum_i \alpha_i = 1, 0 \leq \alpha_i \leq C, \forall i \end{aligned}$$

여기서,  $K(x_i, x_j)$ 는 다음과 같다.

$$\begin{aligned} K(x_i, x_j) &= \phi(x_i) \cdot \phi(x_j) \\ &= \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) & (3) \end{aligned}$$

또한,  $\alpha_i$ 는 라그랑지안 계수이다. KKT condition과 함께 고려할 때,  $\alpha_i = 0$ 인 포인트  $x_i$ 는  $\omega \cdot \phi(x) > \rho$ 의 영역에,  $0 < \alpha_i < C$ 인 포인트  $x_i$ 는  $\omega \cdot \phi(x) = \rho$ 에,  $\alpha_i = C$ 인 포인트  $x_i$ 는  $\omega \cdot \phi(x) < \rho$ 의 영역에 놓여 있다는 것을 의미한다.

## 4. Importance-base FOC-SVM

본 장에서는 [8]의 OC-SVM에 대한 퍼지 버전을 유도한다. 데이터  $x_i$ 의 퍼지 멤버십(fuzzy membership)  $0 < \beta_i \leq 1$ 를 적용한 식은 다음과 같다.

$$\begin{aligned} & \text{minimize} \quad \frac{1}{2} \|\omega\|^2 + C \sum_i \xi_i + \rho & (4) \\ & \text{subject to} \quad \beta_i (1 - \omega \cdot \phi(x)) \leq \rho - \xi_i, \xi_i \geq 0, \forall i \end{aligned}$$

식 (4)를 풀기 위하여  $\alpha_i$ 와  $\gamma_i$ 를 사용한 라그랑지안 계수법을 적용한다.

$$L = \frac{1}{2} \|\omega\|^2 + \rho + C \sum_i \xi_i - \sum_i \alpha_i (\rho - \beta_i (1 - \omega \cdot \phi(x_i)) + \xi_i) - \sum_i \gamma_i \xi_i \quad (5)$$

$$\alpha_i \geq 0, \quad \gamma_i \geq 0$$

$L$ 을  $\omega$ ,  $\rho$ ,  $\xi_i$ 로 각각 미분하여 0으로 놓으면 다음의 식들을 얻을 수 있다.

$$\sum_i \alpha_i = 1 \quad (6)$$

$$\omega = \sum_i \alpha_i \beta_i \phi(x_i) \quad (7)$$

$$\alpha_i = C - \gamma_i \Rightarrow 0 \leq \alpha_i \leq C \quad (8)$$

식 (5)를  $\alpha_i$ 로 다시 표현함으로써 식 (4)의 듀얼 문제인 다음의 식을 얻을 수 있다.

$$\text{maximize} \quad (9)$$

$$W = \sum_j \alpha_i \beta_i - \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j \beta_i \beta_j K(x_i, x_j)$$

KKT condition은 다음과 같다.

$$\gamma_i \xi_i = 0 \quad (10)$$

$$\alpha_i (\rho - \beta_i (1 - \omega \cdot \phi(x_i)) + \xi_i) = 0 \quad (11)$$

### 최적화

식 (9)는 Quadratic Programming을 이용하여 해를 구할 수 있다. 본 논문에서는 SMO (Sequential Minimal Optimization)을 수정하여 이용한다[20]. SMO의 알고리즘은 1) 먼저 2개의 데이터 포인트를 선택하고, 2) 식 (6)과 식 (8)을 만족시키면서 선택된 2개의 데이터 포인트에 해당하는  $\alpha_i$ 를 갱신한다. 3) 모든 데이터 포인트에 대해 KKT조건

을 만족할 때 까지 1)와 2)를 반복 한다[8,20,21].  $\alpha_i$ 에 대한 갱신 식은 다음과 같다.

$$\alpha_2 = \alpha_2^* + \frac{S_1 - S_2 - (\beta_1 - \beta_2)}{\beta_1^2 K_{11} + \beta_2^2 K_{22} - 2\beta_1 \beta_2 K_{12}} \quad (12)$$

$$\alpha_1 = \alpha_1^* + \alpha_2^* - \alpha_2$$

여기서  $S_i$ 는  $\beta_i \phi(x_i)$ 와  $\omega$ 의 내적이다.

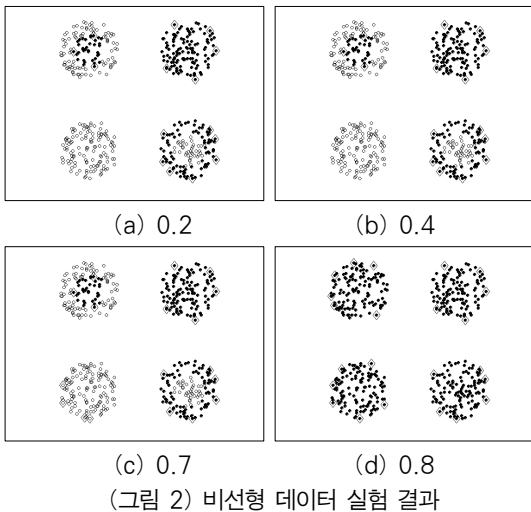
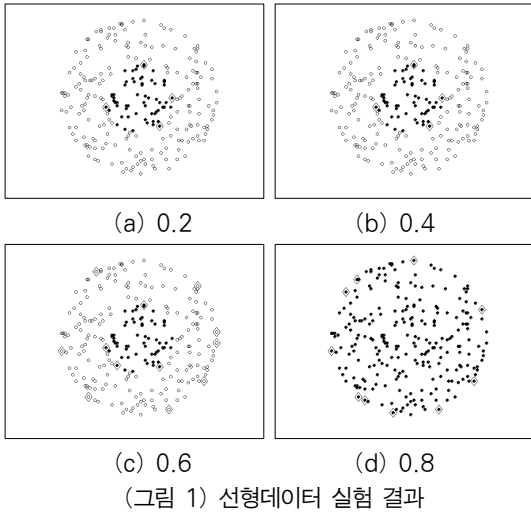
$$S_i = \sum_j \alpha_j \beta_i \beta_j K(x_i, x_j) \quad (13)$$

(그림 1)은 원형을 이루는 가상의 데이터에 대한 IFOC-SVM 수행 결과이다. 이 데이터 샘플은 중심 부분과 외곽 부분으로 분리하여 중요도를 각각 다르게 주었다. 흑점은 0.8이고, 흰 점은 (a) 0.2, (b) 0.4, (c) 0.6이다. 또한, 마름모로 표시된 점은 선정된 서포트 벡터를 의미 한다.

(그림 1)의 (a)와 (b)는 중심 부분의 외곽에서만 서포트 벡터가 표시되어 있다. 이것은 외곽부분의 데이터들에 대한 중심과의 거리 보다 중심부분의 중요도가 크게 높기 때문이다. 그러나 (c)에서는 중요도가 높은 중심 부분에서도, 중심과의 거리가 긴 외곽 부분에서도 서포트 벡터가 표시되어 있다. 이것은 중요도의 거리에 대한 측정치가 비슷하기 때문이다. 마지막으로 (d)는 모든 데이터의 중요도가 같다. 이것은 OC-SVM의 실행 결과와 같은 결과를 보여준다.

(그림 2)는 네 개의 군으로 이루어진 비선형 데이터에 대한 IFOC-SVM의 실험 결과이다. (그림 2)는 비선형 데이터에 대한 실험 결과이지만, (그림 1)에서의 선형데이터의 실행 결과와 비슷한 것을 볼 수 있다. 특히, (c)에서는 중요도의 차가 줄어들면서, 왼쪽 하단의 중요도가 낮은 군에서도 서포트 벡터가 추출되는 것을 볼 수 있다.

(그림 1)과 (그림 2)로부터 IFOC-SVM은 중심과의 거리가 크면서, 중요도가 큰 데이터들을 서포트 벡터로 선정한다는 것을 알 수 있다.



- (2) 분할된 세그먼트를 가지고 각각의 중요도를 (importance measure) 측정하고, 중요도에 대한 퍼지 멤버십 (fuzzy membership)을 구한다.
- (3) 중요도를 가진 세그먼트들을 가지고 IFOC-SVM을 실행하여 결과로 나온 서포트 벡터들을 가지고 비디오 요약을 구축한다.

### 5.1 Importance Measure

IFOC-SVM은 OC-SVM 수행 과정에 사전 정보를 결합할 수 있다. 비디오에서의 사전 정보는 샷의 내용, 샷의 길이, 시각적 특징의 시간적인 변화 등 샷의 다양한 성질로부터 얻을 수 있다. 즉, 세그먼트의 중요도로 사용될 수 있는 다양한 소스들이 존재한다. 더욱이, 이러한 정보들을 직접적으로 다루기에는 애매하거나 모호한 것들이 대부분이다. 다양하고 불확실한 정보들을 결합하고 다루기 위한 방법으로 퍼지 집합 이론을 사용한다. 즉, 비디오 세그먼트의 사람의 주관적인 중요도를 퍼지 멤버십으로 (fuzzy membership) 표현하고, 퍼지 결합 연산자를 (fuzzy aggregation operator) 사용하여 중요도들을 결합한다.

퍼지 멤버십 함수의 집합  $\{f^{(j)}(i), \text{for } i=1\dots n\}$  이 있다고 가정하자. 각  $f^{(j)}(i)$ 는 샷  $j$ 에 대한  $i$  번째 정보 소스로부터 얻어진 중요도에 따른 멤버십이다. 그리고 샷  $j$ 에 대한 결합된 중요도  $S_j$ 는 다음과 같다.

$$S_j = f^{(j)}(1) \oplus f^{(j)}(2) \oplus \dots \oplus f^{(j)}(n) \quad (14)$$

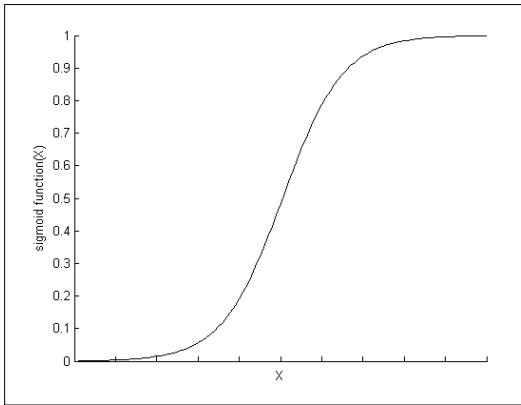
여기서  $\oplus$ 는 결합 연산자 (aggregation operator) 이다[22].

중요도 측정치는 많은 다른 정보 소스로부터 얻을 수 있고, 응용분야에 종속적이다. 본 논문에서는 비디오 요약에 많이 쓰일 수 있는 두 가지 일반적인 접근 방법을 제안한다.

## 5. 비디오 요약

본 논문에서의 비디오 요약은 길이가 긴 비디오를 대상으로 한다. 길이가 긴 비디오를 프레임 단위로 다루기에는 데이터 양이 너무 많고, 비디오 요약도 제대로 이루어 지지 않기 때문에, 비디오를 세그먼트 단위로 분할을 한 후, 최적의 세그먼트 집합을 구한다. 비디오 요약은 3단계로 구성된다.

- (1) 비디오를 세그먼트 단위로 분할한다.



(그림 3) 세그먼트 길이에 따른 멤버십 함수 그래프

**Segment Length** : 비디오 세그먼트의 길이는 중요도로써 가장 쉽게 접근할 수 있는 정보이다. 즉, 시간의 비중이 클수록 더욱 중요한 세그먼트라고 가정 할 수 있다. 세그먼트의 길이를 가지고 중요도 측정에 대한 퍼지 멤버십 함수를  $L(j)$ 라 하고 정의는 다음과 같다.

$$L(j) = 1 / (1 + \exp(-3 \times (x - m) / \sigma)) \quad (15)$$

위 식에서  $x$ 는 세그먼트의 길이,  $m$ 은 세그먼트 길이의 평균,  $\sigma$ 는 세그먼트 길이의 표준편차이다.

(그림 3)은 식 (15)를 그래프로 그린 것이다. 그림에서  $x$ 와 같이 세그먼트의 길이를 비선형적으로  $[0, 1]$ 로 사상한다.

**Segment Content** : 사용자는 비디오 요약에서 특별히 포함 시키거나, 배제 시키고 싶은 세그먼트들이 존재 할 수 있다. 이러한 사용자의 요구를 충족시킬 수 있는 “Query by Example” 형식을 사용한다. 즉, 사용자는 비디오 요약 시스템에게 “positive example” 또는 “negative example”을 주고, 시스템은 사용자의 입력에 따라 중요도의 퍼지 멤버십을 계산 할 수 있다. positive sample을  $x_p$ , negative sample을  $x_n$ 이라 하자. 세그먼트  $j$ 의 특징 벡터는  $x_j$ 이다.  $x_p$ 과  $x_j$ 의 관계 정도를 나타

내는  $P(j)$ 와  $x_n$ 과  $x_j$ 의 관계 정도를 나타내는  $N(j)$ 를 다음과 같이 정의 한다.

$$P(j) = K(x_p, x_j) \quad (16)$$

$$N(j) = 1 - K(x_p, x_j) \quad (17)$$

$K(\cdot, \cdot)$ 는 가우시안 커널 공간에서의 내적을 나타낸다.

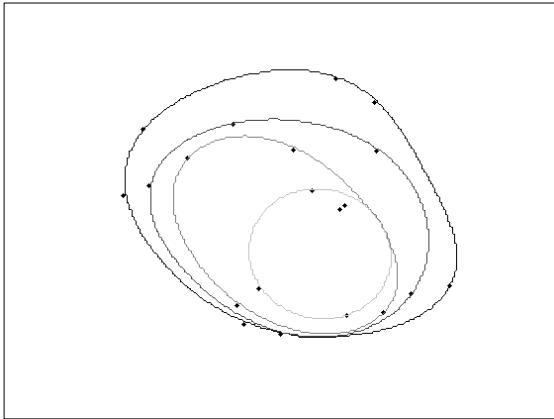
식 (14)를 이용하여 세그먼트의 길이에 따른 중요도와 세그먼트 내용에 따른 중요도를 결합할 수 있다. 본 논문에서는 min/max 연산자를 사용하여 결합된 중요도 멤버십을 계산할 수 있다.

$$\beta_j = \min(\max(L(j), P(j)), N(j)) \quad (18)$$

## 5.2 확장가능 비디오 요약

IFOC-SVM은 파라미터  $C$ 의 값을 다양하게 조절함으로써 아웃라이어와 (outlier) 서포트 벡터의 비율을 제어할 수 있다. 특히,  $C$ 가 1일 때는 아웃라이어 없이 모든 데이터를 포함하는 최소구의 표면에 있는 최적의 서포트 벡터들을 찾는다. 확장가능 비디오 요약 알고리즘은  $C$ 를 1로 설정하여 IFOC-SVM을 실행 한다. 실행 결과로써 얻어진 서포트 벡터들은 키프레임 집합이 된다. 선정된 서포트 벡터들은 데이터 제거되고, 남아있는 데이터를 가지고 IFOC-SVM을 수행해 새로운 서포트 벡터들을 찾는다. 즉, 구의 표면을 하나씩 벗겨내면서 전보다는 작지만 새로운 최소 구를 찾는다. 새로운 최소 구로부터 얻어진 서포트 벡터들은 이미 얻어진 서포트 벡터 집합에 추가되어져서 계층적 단계를 이루게 된다[23].

(그림 4)는 IFOC-SVM을 사용하여 확장가능 비디오 요약기법을 나타내고 있다. 각 단계 별로 IFOC-SVM을 수행하여 선정된 데이터 포인트들을 contour 형식으로 연결하여 보여주고 있다.



(그림 4) 확장 가능 비디오 요약 기법

## 6. 실험

본 논문에서는 제안한 비디오 요약 알고리즘을 50분 비디오, 5분 뮤직비디오, 10분의 다큐멘터리 비디오에 대하여 실험을 하였다. 시각적 특징 벡터로써  $6 \times 6 \times 6$  RGB 칼라 히스토그램을 (color histogram) 사용하였다. 또한, 비디오를 세그먼트로 분할하기 위하여 계층적 클러스터링을 (hierarchical clustering) 적용하였고, 세그먼트의 키프레임으로써 세그먼트의 중심에 가장 가까운 프레임을 키프레임으로 선정하였다. 또한, 비디오 요약을 단계별로 점점 상세히 볼 수 있는 scalable video summary scheme을 사용하였다.

### 6.1 세그먼트의 길이를 반영한 비디오 요약

제안한 비디오 요약 알고리즘의 성능을 알아보기 위하여, 우선 세그먼트의 길이를 중요도 측정치로 사용하였다. 세그먼트의 길이를 IFOC-SVM의 멤버십으로 변환하기 위하여 식 (15)에서와 같이 시그모이드(sigmoid) 함수를 사용하였다.

(그림 5)와 (그림 6)은 세그먼트 길이를 IFOC-SVM의 멤버십으로 적용하여 실험한 결과이다. (그림 5)와 (그림 6)에서의 그림들은 IFOC-SVM의 결과인 서포트 벡터들로 선정된 키프레임의 정지영상들 (image) 이다. 또한, (a)와 (b)를 통해

서 각 단계별로 진행 할수록 비디오 요약이 점점 상세해 지는 것을 볼 수 있다. 이것은 IFOC-SVM의 “scalable” 성질을 이용한 비디오 요약 형식이다.

### 6.2 사용자의 성향을 반영한 비디오 요약

사용자의 성향이 반영 되었는지를 확인하기 위하여 실험 1과 같은 비디오 데이터를 가지고 실험하였다.

(그림 7)과 (그림 8)에서 (a)는 사용자가 꼭 선정되었으면 하는 데이터 샘플이고, (b)는 사용자가 비디오 요약의 결과에 선정되지 말았으면 하는 데이터 샘플이다. 실험 1에서의 결과인 (그림 5)과 (그림 6)을 (그림 7)과 (그림 8)에 대해서 비교하면, negative 샘플과 비슷한 키프레임은 실험 1에는 있었지만, 실험 2의 결과에는 볼 수 없다. 또한, positive 샘플과 비슷한 키프레임들은 실험 1에는 없었지만, 실험 2에서는 사용자의 성향이 반영되어 서포트 벡터로 선정된 것을 볼 수 있다.

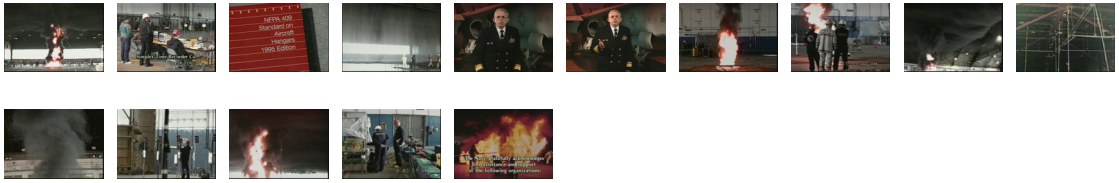
위 실험 등을 통하여 세그먼트의 길이, 사용자 성향 이외에도 다양한 비디오의 정보들을 반영하여 비디오 요약에 적용할 수 있다는 것을 알 수 있다.

### 6.3 비디오 요약 성능 비교

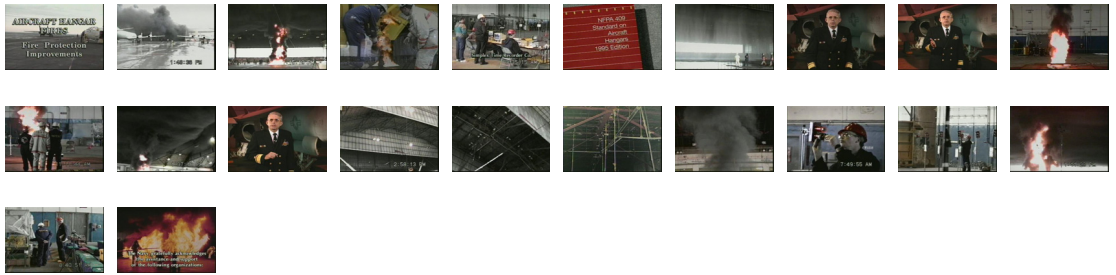
비디오 요약 성능 기준은 주관적인 경우가 많아 다른 연구들과 성능을 비교 평가하기 어렵다 [2,3,5,15]. 그러나 [3]에서는 객관적인 성능 측정 기준으로 이벤트 장악력(event coverage)과 길이 장악력(duration coverage)을 제안 하였다. 이벤트 장악력은 선정된 비디오 세그먼트들이 얼마나 많은 비디오 이벤트에 나타났는지를 측정하는 것이고, 길이 장악력은 선정된 비디오 세그먼트들의 전체 길이를 계산한 것이다.

그러나 이들 측정 기준들은 비디오 요약의 “시각적인 다양성”을 반영하지는 못한다. 만일 이벤트를 이루고 있는 몇 개의 세그먼트들 중 하나를 선택해야 한다면, 길이가 길면서 많이 다른 세그



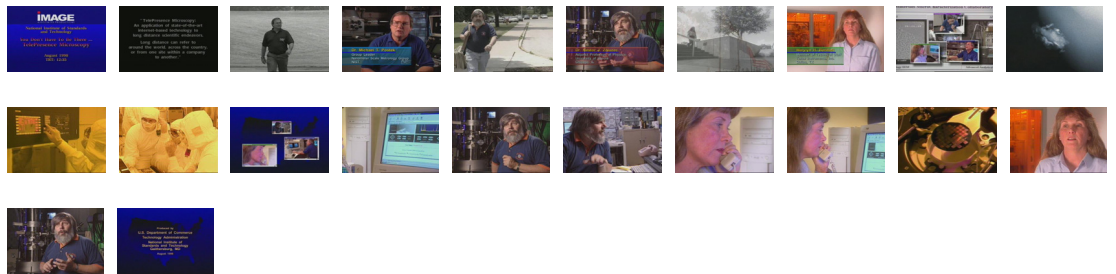


(a) level 1



(b) level 2

(그림 5) 세그먼트 길이를 반영한 비디오 요약 결과

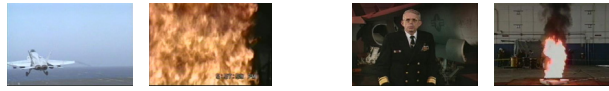


(a) level 1



(b) level 2

(그림 6) 세그먼트 길이를 반영한 비디오 요약 결과

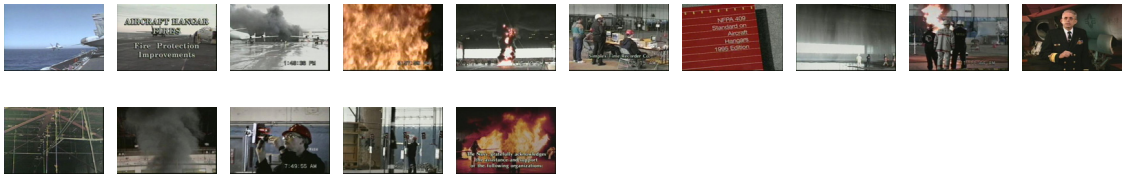


(a) positive

(b) negative



(c) level 1



(d) level 2

(그림 7) 세그먼트 길이와 사용자의 성향을 반영한 비디오 요약 결과

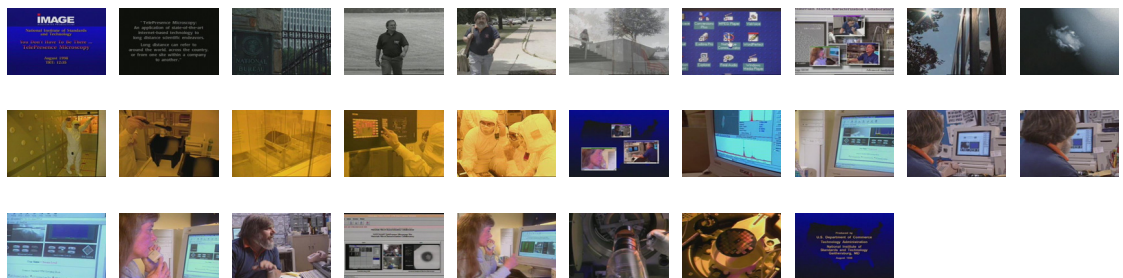


(a) positive

(b) negative

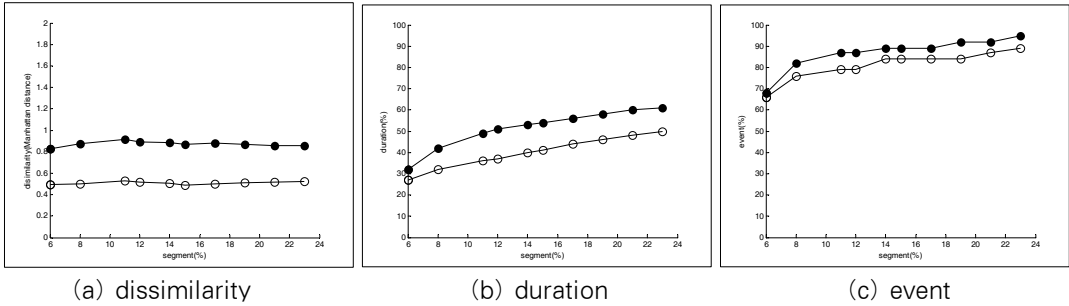


(c) level 1

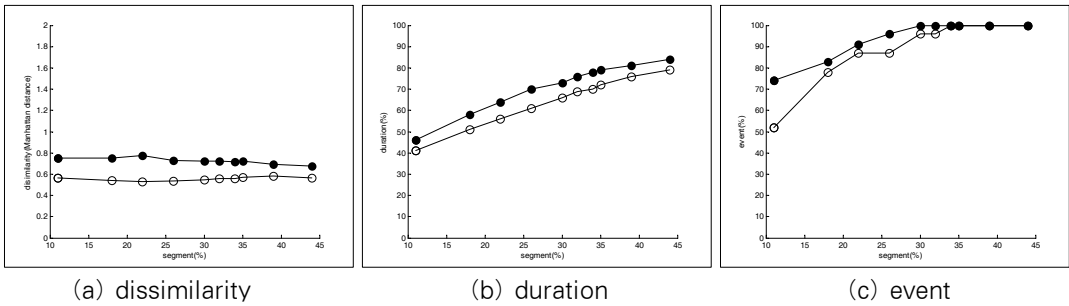


(d) level 2

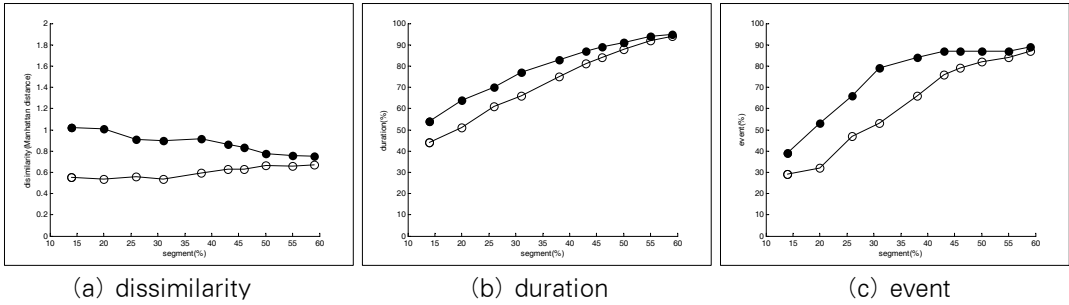
(그림 8) 세그먼트 길이와 사용자의 성향을 반영한 비디오 요약 결과



(그림 9) 영화로 실험한 성능 비교 그래프



(그림 10) 뮤직비디오로 실험한 성능 비교 그래프



(그림 11) NIST 데이터로 실험한 성능 비교 그래프

먼트들을 선정하는 것이 좋을 것이다. 이러한 관점으로 비디오 요약의 성능을 측정하기 위하여, 본 논문에서는 “self-dissimilarity”를 정의 하였다. 식 (19)에서  $S$ 는 선정된 비디오 세그먼트의 집합,  $n$ 은  $S$ 의 개수를 나타낸다. 두 개의 세그먼트의 거리를 측정하기 위하여  $L1$ -distance를 사용하였다.

$$\sum_{i \in S} \sum_{j \in S} \|x_i - x_j\| / n^2 \quad (19)$$

(그림 9, 10, 11)은 각각 영화, 뮤직비디오, NIST의 다큐멘터리 데이터를 가지고 [3]의 Manga와의 성능을 이벤트 장악력, 길이 장악력, 시각적 다양성 관점으로 비교한 그래프이다. 각 그림에서 흑점은 제안하는 방법의 성능을 나타내고 백점은 Manga의 성능을 나타낸다. 그래프에서 보면 이벤트 장악력과 길이 장악력에서는 제안한 알고리즘이 Manga 보다 성능 그래프가 조금 높은 것을 볼 수 있다. 그러나 시각적 다양

성에서는 제안한 알고리즘이 Manga 보다 눈에 띄게 우수한 것을 볼 수 있다. 이것은 제안한 비디오 요약의 두 가지 기준에서 “시각적으로 특징 있는” 세그먼트를 선정하는 것이 Manga 보다 좋다는 것을 보여준다.

## 7. 결 론

본 논문에서는 비디오 요약을 시각적으로 특징이 있고, 주관적으로 중요한 세그먼트를 최적으로 선정하는 것으로 새롭게 해석 하였다. 이러한 비디오 요약을 위하여 비디오 세그먼트의 다양한 정보들에 대한 중요도를 측정하고 이를 비디오 요약에 적용하였다. 이를 위하여, 기존의 OC-SVM에 주관적인 중요도를 결합한 IFOC-SVM을 유도 하였다.

또한, 본 논문에서는 비디오 요약의 성능을 비교 분석하기 위한 새로운 성능 측정 기준을 제시 하였다. 이것은 서로 비교할 수 없었던 기존의 다양한 비디오 요약 알고리즘들에 대한 새로운 성능 기준이 될 수 있다. 마지막으로 본 논문에서 제안한 비디오 요약의 성능을 증명하기 위하여 가상의 데이터와 여러 형태의 비디오들을 가지고 실험을 하였다. 실험 결과는 제안하는 비디오 요약 알고리즘이 이벤트 장악력, 길이 장악력과 시각적 다양성 면에서 Manga 보다 성능이 우수함을 보여주었다.

## 참 고 문 헌

- [1] Nevenka Dimitrova, et al. “Applications of Video-Content Analysis and Retrieval”, IEEE Multimedia, Vol. 9 Issue. 3, Jul-Sep 2002, pp.42-55
- [2] Yeung, M.M., Boon-Lock Yeo: Video Visualization for Compact Presentation and Fast Browsing of Pictorial Content. IEEE Transactions on Circuits and Systems for Video Technology. Vol.7 No.5 October (1997) 771-785
- [3] Shingo Uchihashi, Jonathan Foote, Andreas Girgensohn, John Boreczky: Video Magna: Generating Semantically Meaningful Video Summaries. Proceedings of ACM International Conference on Multimedia (1999) 383--391
- [4] Dimitrova, N., Hong-Jiang Zhang, Shahraray, B., Sezan, I., Huang, T., Zakhor, A.: Applications of Video-Content Analysis and Retrieval. IEEE Multimedia Vol. 9 (2002) 42--55
- [5] Girgensohn, A., Boreczky, J., Wilcox, L.: Keyframe-Based User Interfaces for Digital Video. IEEE Computer September (2001) 61--67
- [6] Muller, K.-R., Mika, S., Ratsch, G., Tsuda, K., Scholkopf, B.: An Introduction to Kernel-Based Learning Algorithms. IEEE Transactions on Neural Networks Vol.12 No.2 March (2001) 181--202
- [7] Yunqiang Chen, Xiang Sean Zhou, Huang, T.S.: One-class SVM for learning in image retrieval. Proceedings of International Conference on Image Processing Vol.1 (2001) 34--37.
- [8] B. Scholkopf, J.C. Platt, J. Shawe-Taylor, A. J. Smola, R. C. Williamson: Estimating the Support of a High-Dimensional Distribution. Microsoft Research Technical Report MSR-TR-99-87 (1999)
- [9] Pei-Yi Hao “Fuzzy one-class support vector machines” Fuzzy Sets and Systems 159 (2008) 2317 - . 2336
- [10] 김기주, 최영식 “퍼지 원 클래스 서포트 벡터 머신”, 인터넷정보학회논문지 제6권 32호, 2005
- [11] YoungSik Choi and KiJoo Kim “Video

- Summarization Using Fuzzy One-Class Support Vector Machine” LNCS3043, pp.49-56, 2004
- [12] S. W. Smoliar and H. J. Zhang, “Content-based video indexing and retrieval”, IEEE Multimedia, 1994, pp. 62-72.
- [13] B. Shahraray and D. C. Gibbon, “Automatic generation of pictorial transcripts of video program,s”, in Proc. IS&T/SPIE Digital Video Compression: Algorithms and Technologies, SanJose,CA,1995,pp.512-519.
- [14] Y. Zhuang, Y. Rui, T. S. Huang, and S. Mehrotra, “Adaptive key frame extraction using unsupervised clustering”, ICIP’98, vol. 1, Oct 1998, pp.866-870.
- [15] Hanjalic, A., HongJiang Zhang: An Integrated Scheme for Automated Video Abstraction Based on Unsupervised Cluster-Validity Analysis. IEEE Transactions on Circuits and Systems for Video Technology Vol.9 No.8 December (1999) 1280--1289
- [16] Asa Ben-Hur, David Horn, Hava T. Siegelmann, and Vladimir Vapnik, “Support Vector Clustering”, Journal of Machine Learning Research 2, pp.125-137, 2001.
- [17] Manuel Davy and simon Godsill, “Detection of abrupt spectral changes using support vector machines: an application to audio signal segmentation”, IEEE International Conference on Acoustics, Speech, and signal Processing, Vol. 2, 12-17, pp.1313-1316, May 2002.
- [18] P. Hayton, B. Scholkopf, L. Tarassenko, and P. Anuzis, “Support vector novelty detection applied to jet engine vibration spectra”, in NIPS’2000, 2000.
- [19] Tax, D. M. J. and Duin, R. P. W., “Image database retrieval with Support vector data description”, Proceedings of the Sixth Annual Conference of the Advanced School for Computing and Imaging, ASCI, Delft, June 2000.
- [20] J. C. Platt “Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines”, Microsoft Research Technical Report, MST-TR-98-14, 1998.
- [21] Nello Cristianini and John Shawe-Taylor, “An Introduction to Support Vector Machines and other kernel-based learning methos”, Cambridge University Press, 2000.
- [22] J. M. Keller, et. al., “Evidence Aggregation Networks for Fuzzy Logic Inference”, IEEE Transactions on Neural Networks, Vol. 3, No. 5, pp. 762-769, 1992
- [23] YoungSik Choi, Sang-Yoon Lee: Scalable Keyframe Extraction Using One-Class Support Vector Machine. ICCS2003 LNCS2660 (2003) 491-499

## ● 저 자 소 개 ●

### 김 기 주



2004년 한국항공대학교 컴퓨터공학과 졸업(학사)  
2006년 한국항공대학교 대학원 컴퓨터공학과 졸업(석사)  
2006년~현재 한국항공대학교 대학원 컴퓨터공학과 박사과정  
관심분야 : 정보검색, 데이터마이닝, 패턴인식, etc.  
E-mail : zgeniez@kau.ac.kr

### 최 영 식



1985년 연세대학교 전자공학과 졸업(학사)  
1987년 연세대학교 대학원 전자공학과 졸업(석사)  
1996년 미국 Missouri-Columbia대학교 대학원 Computer Engineering학과 졸업(박사)  
2002년~현재 한국항공대대학 항공전자및정보통신공학부 교수  
관심분야 : 클라우드 컴퓨팅, 정보검색, 데이터마이닝, 영상처리, etc.  
E-mail : choimail@kau.ac.kr