

논문 2015-52-1-12

# 정합 쌍의 통계적 분석을 이용한 정형/비정형 객체 영상의 적응적 정합 방법

( Adaptive Matching Method of Rigid and Deformable Object Image  
using Statistical Analysis of Matching-pairs )

원 인 수\*, 양 훈 준\*, 장 혁\*, 정 동 식\*\*\*

( In-su Won<sup>Ⓞ</sup>, Hun-Jun Yang, Hyeok Jang, and Dong-Seok Jeong )

## 요 약

본 논문은 동일한 특징을 사용하여 정형 객체와 비정형 객체 영상들을 정합할 수 있는 적응형 정합 방법을 제안한다. 이를 위한 방법으로 우선 기하학적 검증으로 두 영상의 정합 여부를 결정하고 정합 정보를 생성한다. 그리고 정합 정보의 통계적 분석을 통해 비정형 정합 쌍과 비정합 정합 쌍을 분류하는 결정 경계를 구한다. 제안된 방법의 성능 평가 결과는 기존의 방법과 비교하였을 때, 복잡도는 낮았으며, 정합 성공률과 정확도는 높아짐을 보여주었다.

## Abstract

In this paper, adaptive matching method using the same features for rigid and deformable object images is proposed. Firstly, we determine whether the two images are matched or not using the geometric verification and generate the matching information. Decision boundary which separates deformable matching-pair from non-matching pair is obtained through statistical analysis of matching information. The experimental result shows that the proposed method lowers the computational complexity and increases the matching accuracy compared to the existing method.

**Keywords** : Rigid image matching, Deformable image matching, Statistic approach.

## I. 서 론

최근 스마트 기기의 급속한 보급과 광대해진 빅데이터의 활용을 위해 컴퓨터 비전은 다양한 산업 분야로 파급되어 사용되고 있다. 스마트 기기의 카메라를 이용

한 얼굴 인식, 동작 인식, 증강 현실 등은 실생활에서 활용되고 있으며, 검색 등은 전통적인 문자 기반 검색 방법에서 영상 검색 기반으로 발전되고 있다. 이를 위해 MPEG-7 에서는 정지영상에 대한 내용기반 검색을 위한 방법을 표준으로 재정하였으며, 최근에는 MPEG-7 Compact Descriptor Visual Search(CDVS)<sup>[1]</sup>를 구성하여 모바일 기반에서 영상을 빠르게 검색할 수 있는 연구가 진행 중에 있다. 컴퓨터 비전의 주요한 알고리즘은 크게 영상의 특징 추출 방법과 특징 정합 방법으로 구분된다. 대표적인 특징 추출 방법은 크기와 각도 변화에 강인한 D. Lowe의 SIFT(Scale-Invariant Feature Transform)<sup>[2]</sup>, 영역을 기반으로 하는 J. Matas의 MSER(Maximally Stable Extremal Regions)<sup>[3]</sup> 그리

\* 학생회원, \*\* 정회원, 인하대학교 전자공학과  
(Dept. of Electronic Engineering, Inha University)

Ⓞ Corresponding Author (E-mail: woninsu@inha.edu)

※ 본 연구는 미래창조과학부 및 정보통신기술진흥센터의 정보통신·방송 연구개발사업의 일환으로 수행하였음. [2014(R2012030111), UVD(Unified Visual Descriptor) 기반 Smart Mobile Search 기술 개발]  
접수일자: 2014년07월29일, 수정일자: 2014년11월12일  
게재확정: 2014년12월26일

고, 어파인 변화에 강인한 K. Mikolajczyk의 Harris affine detector<sup>[4]</sup> 등이 있다. 그러나 대부분의 영상 특징 추출 및 정합 방법은 영상 식별(identification)에 중점을 두고 있다. 영상 식별이란 형태가 변하지 않는 정형 객체나 건물을 대상으로 시점, 크기, 각도 및 밝기 변화 등에도 강인한 특징을 추출하고 정합하여 정확히 같은 영상을 찾아주는 기법으로 다양한 환경에 의해 변형된 영상에서 질의 영상을 찾아주는 것을 목적으로 하고 있다. 이와 비슷한 개념인 영상 검색(retrieval)은 색상, 질감, 모양 등 하위 수준의 특징들을 사용하여 질의 영상과 유사한 영상을 찾아주는 기법으로 비슷한 색상, 질감, 모양 등을 가진 영상, 동물과 식물 같은 모양이 유사한 객체 등을 분류하는 목적으로 한다. 이렇듯 컴퓨터 비전 기술은 목적에 따라 크게 검색과 식별 기술로 구분되지만, 비정형 객체(deformable object) 영상을 대상으로 하는 연구는 새로운 논점이 되고 있다. 비정형 객체 영상의 정합에 대한 연구는 객체 인식, 3D 영상 생성, 영상 분할 등의 분야에서 진행되고 있으며, 최근 의류나 가방 등과 같은 객체지만 모양이 다양하게 변할 수 있는 같은 비정형 객체 영상을 정합하거나, 동식물들의 분류를 위한 정합 기술이 새로운 도전 과제로 등장하고 있다. 그러나 정형 객체 영상 정합 방법과 비정형 객체 영상 정합 방법은 목적 자체가 다르기 때문에 특징 추출, 정합 등의 방법을 동일하지 않다. 따라서 두 방법을 동시에 사용하기에는 많은 어려움이 있다.

본 논문은 동일한 특징을 사용하여 정형 객체 영상과 비정형 객체 영상을 정합할 수 있는 단일화 방법을 제안한다. 제안된 방법은 입력 영상에서 동일한 특징점을 추출하며 두 개의 정합 쌍을 동시에 구성한다. 그리고 기하학적 검증(Geometry Verification)을 수행하여 정형 정합 여부를 판단하게 된다. 검증 결과로 정합 쌍의 개수, 올바른 정합 쌍의 개수 등의 정합 정보가 구성되고 이를 이용하여 정형 객체 영상들의 정합 여부를 판단할 수 있다. 기존의 방법은 검증을 통과하지 못한 경우는 비정합으로 결정이 되지만, 본 논문은 정합 정보를 이용하여 비정형 객체 영상들의 정합 쌍일 가능성이 높은 경우를 통계적 분석을 이용하여 찾아내었다. 이를 위해 비정형 객체 영상들과 비정합 객체 영상들을 훈련 영상으로 구성하고 각 영상에서 정합 정보를 계산한다. 그리고 정합 정보들의 통계적 분석을 통해 비정형 객체 영상의 정합 쌍일 확률이 높은 경우 비정형 객체 영상

정합 방법을 수행한다.

본 논문의 구성은 다음과 같다. 제안된 정형과 비정형 객체 영상의 정합 방법에 대해 II장에서 소개하고, III장에서는 실험을 통해 제안된 방법의 성능을 검증하며, IV장에서 결론을 맺는다.

## II. 정형/비정형 객체 영상 정합 방법

제안된 정형/비정형 객체 영상 정합 방법의 전체 순서도는 그림 1과 같다. 질의 영상과 비교 영상에서 특징 검출과 서술자 추출을 하고 특징점들의 서술자 차이를 비교하여 1:1 정합 쌍과 1:N 정합 쌍을 구성한다. 정형 정합 여부를 판단하기 위해 1:1 정합 쌍을 이용하여 기하학적 검증을 수행하고 정합 정보를 생성한다. 정형 정합으로 결정되지 않으면 정합 정보를 이용하여 비정형 정합 쌍일 가능성이 높은 경우를 구분한다. 이를 위해 훈련 셋을 이용하여 비정형과 비정합을 구분할 수 있는 분류법을 사용하고 비정형의 가능성이 높은 영상은 1:N 정합 쌍을 이용하여 비정형 정합 방법을 수행한다. 비정형 정합 방법은 각 정합 쌍들의 기하학적 유사

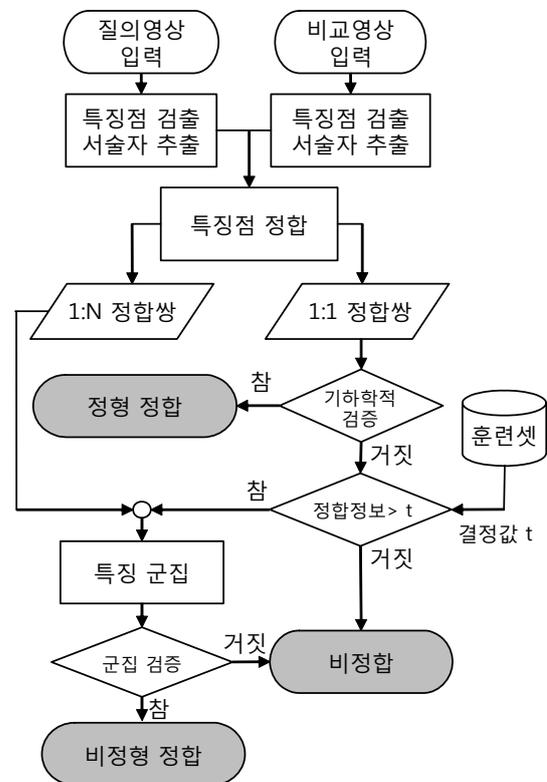


그림 1. 제안된 영상 정합 시스템의 순서도  
Fig. 1. Flowchart of proposed image matching system.

성을 이용하여 군집을 구성하고, 최종적으로 하나 이상의 군집이 구성되면 비정형 정합이라고 판단한다.

### 1. 특징점 추출

제안된 정합 방법을 위해서 영상의 특징으로 영상의 좌표 값과 유사한 특징들을 군집하기 위한 속성이 필요하다. 특징 군집을 위해 특징의 크기, 주방향성 등을 포함하는 SIFT를 사용하였다. 이와 유사한 특징을 갖는 MSER, Affine detector도 사용 가능하다. SIFT는 특징점의 위치, 크기 및 방향을 결정하기 위해 Difference of Gaussian(DoG)를 사용하여 다른 크기 공간(scale space)에서 극점(extrema)를 찾는 방법으로, 영상의 어파인(affine) 변환이나 투시 변환에 대하여도 특징의 반복성이 높다. SIFT의 서술자는 검출된 특징점을 중심으로 크기 영역을 주방향성을 기준으로 4×4으로 구분하고, 각 영역에서 주변 화소의 차이를 8 방향성으로 표현하여 128차원의 히스토그램을 구성한다. 각 영상에서 추출되는 특징은 식 1과 같이 나타낼 수 있다.

$$F(i) = \{c_i, s_i, o_i, d_i\}, i = 0 \sim N. \quad (1)$$

$c_i$ 는 좌표값,  $s_i$ 는 크기,  $o_i$ 는 주방향,  $d_i$ 는 128차원의 서술자를 나타낸다.

### 2. 특징점 정합

두 개의 영상 간에 대응되는 점, 즉 정합 쌍을 찾는 과정을 특징점 정합이라 한다. 영상에서 각 특징점들의 서술자 차이를 계산하고 차이가 작은 것이 같은 특징점일 가능성이 높다고 가정한다. 정합 쌍을 구성하는 방법에 따라 1:1 정합 쌍과, 1:N 정합 쌍으로 구분한다. 1:1 정합 쌍은 일반적으로 정형 객체 영상의 정합 방법에서 사용한다. 1:1 정합 쌍은 특징점들의 차이 값이 가

장 작은 정합 쌍과 두 번째로 작은 정합 쌍과의 비율과 비교하는 방법으로 특정 정합 쌍의 분별력이 높은 장점이 있다. 1:N 정합 쌍은 비정형 객체 영상의 정합을 위해 구성하며 차이가 특정 임계값보다 작은 것은 모두 가능성을 가진 정합 쌍이라고 판단하여 최대 N개까지 정합 쌍을 구성하게 된다. 그림 2는 구성된 특징점 정합 쌍의 예시를 나타낸다.

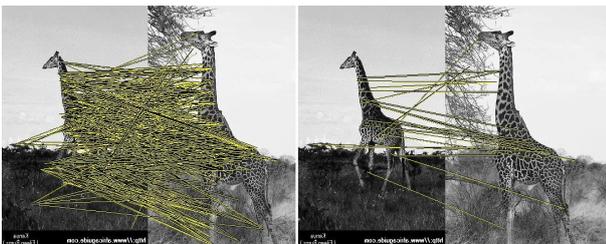
### 3. 기하학적 검증을 통한 정형 정합 결정

정합 쌍이 구성되면 두 영상의 정형 정합을 결정하기 위해 1:1 정합 쌍으로 기하학적 검증을 통해 영상 정합 여부를 결정한다. 대표적인 기하학적 검증 방법으로 RANSAC<sup>[5]</sup> 등이 있다. 그러나 이러한 방법들은 올바른 정합 쌍(inlier)에만 중점을 두고 있다. S. Lepsoy 등은 틀린 정합 쌍(outlier)의 로그 거리비 (Log distance Ratio, LDR)는 올바른 정합 쌍의 특성과 분명히 다를 수 있으므로 정합 쌍의 히스토그램과 틀린 정합 쌍의 확률 모델 간의 연관성을 판단하여 정합 여부를 판단하는 DISTRAT(Distance Ratio Statistic)<sup>[6]</sup>을 제안하였다. 이 방법은 RANSAC 등과 비교하여 성능은 거의 유사하지만 속도가 빠르다는 장점을 갖고 있다.

본 논문에서는 정형 영상의 정합을 위해 DISTRAT을 사용한다. 특징점 정합에서 구성된 1:1 정합쌍을  $(x_1, y_1), \dots, (x_N, y_N)$ 이라 하였을 때, 로그 거리비는 다음 식 2와 같이 표현된다.

$$z_{ij} = \ln\left(\frac{\|x_i - x_j\|}{\|y_i - y_j\|}\right), Z = \{z_{ij} | i \neq j\}. \quad (2)$$

$x_{ij}$ 은 첫 번째 영상에서의 특징점 좌표,  $y_{ij}$ 은  $x_{ij}$ 과 정합된 특징점의 좌표를 의미한다. 로그 거리비를 히스토그램  $h(k)$ 으로 표현하고 전체 정합 쌍을 확률 분포 함수  $f(k)$ ,  $k = 1, \dots, k$ 로 나타내었을 때, 올바른 정합 쌍의 히스토그램과 틀린 정합 쌍은 서로 다른 모양을 갖게 되는데, 이를 이용하여 두 영상의 정합 여부를 판단할 수 있다.  $h(k)$ 와  $f(k)$ 를 비교하기 위해 피어슨의  $\chi^2$  검정<sup>[7]</sup>을 사용하며  $h(k)$ 와  $f(k)$ 가 유사한 모양을 보이면 틀린 영상 정합으로 판단하고 다른 모양을 보이면 정합 영상일 가능성이 높다고 판단한다. 식 3은 이런 유사성을 척도로 계산하는 수식이다.  $c$ 의 값이 커질수록  $h(k)$ 와  $f(k)$ 가 많은 차이가 있음을 나타내며,



(a) 1:N 정합 쌍 (b) 1:1 정합 쌍

그림 2. 특징점 정합 쌍의 예시

Fig. 2. Example of feature matching-pairs.

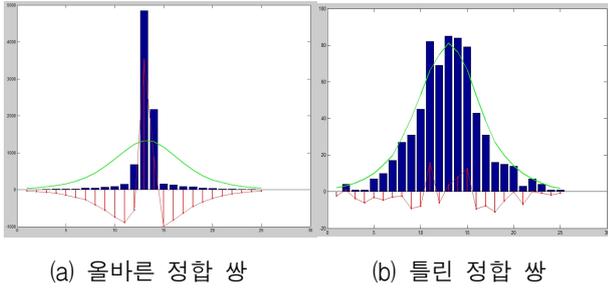


그림 3. 정합 쌍의 LDR 히스토그램  $h(k)$ , 확률 모델 함수  $f(k)$ , 그리고 차이값  $d(k)$ 의 예제  
Fig. 3. Example of LDR histogram  $h(k)$ , model function  $f(k)$ , and different  $d(k)$ .

이는 정합쌍이 올바른 정합 쌍을 많이 가지고 있다는 것을 의미한다.

$$c = \sum_{k=1}^K \frac{(h^k - n f_k)^2}{n f_k} \geq \chi_{1-\alpha, K-1}^2 \quad (3)$$

$n$  은 히스토그램을 구성하는 전체 정합 쌍의 개수이고,  $\chi_{1-\alpha, K-1}^2$  는  $K-1$  의 자유도를 갖는  $\chi^2$  분포의  $100 \cdot (1-\alpha)$  의 백분율을 정의하는 임계값이다. 그림 3은 정합과 비정합 영상의 정합 쌍에서 각각  $h(k)$ ,  $f(k)$ 와  $d(k)$ 를 도식화한 예제이다.

올바른 정합 쌍의 수를 예측하기 위해서  $h(k)$ 와  $f(k)$ 의 차이값인  $d(k)$ 의 고유치를 이용한다. 식 4는  $d(k)$ 를 계산하는 수식이며,  $\beta$  는  $h(k)$ 와  $f(k)$ 의 비율을 맞추주기 위한 가중치이다.

$$d(k) = h(k) - \beta f(k), \quad \beta = \frac{\sum_{k=1}^K h(k)f(k)}{\sum_{k=1}^K (f(k))^2} \quad (4)$$

$d(k)$ 와 로그 거리비를 이용하여 올바른 정합 쌍을 확인하기 위한 행렬  $D$ 를 다음 식 5와 같이 계산한다.

$$D_{ij} = \begin{cases} d_q(z_{ij}) & i \neq j \\ 0 & i = j \end{cases} \quad (5)$$

여기서  $d_q$  는  $d(k)$ 를 양자화한 값이며,  $D$ 로부터 고유치  $\mu$  와 우세한 고유벡터  $r$  을  $Dr = \mu r$  로 구한다. 계산된 고유치와 차이값을 이용하여 다음 식 6을 이용하여 올바른 정합 쌍의 개수를 예측할 수 있다. 또한, 고유벡터  $r$ 을 내림차순으로 정렬하고, 예측된 정합 쌍의 개수만큼의 상위에 해당하는 정합 쌍들을 최종적으로 올바른 정합 쌍으로 판단한다.

$$inlier_n = 1 + \frac{\mu}{\max_{k=1, \dots, K} d(k)} \quad (6)$$

마지막으로 정합된 쌍들과 올바른 정합 쌍들의 가중치를 계산하고, 가중치의 비율을 이용하여 정합 여부를 판단한다. 정합 쌍들의 가중치는 특징점 정합시 가장 가까운 거리 값과 두 번째로 가까운 거리 값의 비율이 주어졌을 때, 올바른 정합일 조건부 확률을 사용한다. 식 7은 이에 대한 조건부 확률식이다.

$$p(c|r) = \frac{p(r|c)}{p(r|c) + p(r|\bar{c})} \quad (7)$$

여기서  $c$ 와  $\bar{c}$ 는 올바른 정합쌍일 확률과 틀린 정합을 의미하며,  $r$  은 거리비이다.

#### 4. 정합 쌍의 통계적 분석을 이용한 비정합 정합 쌍과 비정형 정합 쌍의 분리 방법

두 영상  $x, y$  의 정형 정합 결정 방법의 결과로 정합 쌍의 개수( $match_n$ ), 올바른 정합 쌍의 개수( $inlier_n$ ), 정합 쌍의 가중치( $weight$ )를 알 수 있으며, 이를 정합 정보  $M(x, y)$  라 한다. 정합 정보는 다음 식 8 과 같이 구성된다.

$$M(x, y) = \{match_n, inlier_n, weight\} \quad (8)$$

$M(x, y)$ 를 이용하여 정형 정합이 되지 않은 영상들에 대해 비정형 정합과 비정합 정합을 구분한다. 이를 구분하는 이유는 비정형 객체 영상의 정합 방법은 정합 성능은 좋지만, 복잡도가 높기 때문에 비정합일 가능성이 높은 영상들에 대해서는 우선적으로 제외하여 복잡도를 낮추기 위함이다.

##### 가. 비정형/비정형 객체 영상의 통계적 분석

비정형 객체 영상과 비정합 객체 영상의 구분을 위해 훈련 영상에서  $M(x, y)$ 를 구성한다.  $M(x, y)$ 의 통계적 분석을 위해 각 속성들의 평균  $m$  와 분산  $\sigma^2$ 을 계산하여 정규 분포 모델  $N(m, \sigma^2)$ 을 계산할 수 있다. 그러나 비정형과 비정합의 정합 정보는 많은 부분에서 유사성을 보이기 때문에 두 부류를 구분하기 어려움이 있다. 그림 4는 훈련 셋에서의 비정합과 비정형 정합 쌍들의 각 정보를 정규 분포를 비교한 그림이다. 그러나 그림에서도 나타나듯 각 정보들의 정규 분포 모델로는

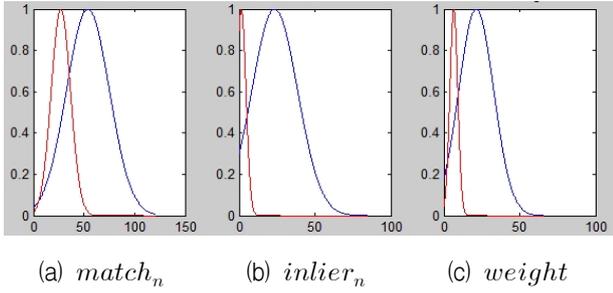


그림 4. 비정형과 비정합 정합 쌍에 대한 각 정합 정보의 정규 분포 모델

Fig. 4. Gaussian distribution model of non-rigid and non-matching-pair information.

베이시안 분류기로 두 부류를 구분하기에는 오류 확률이 높음을 알 수 있다. 오류 확률을 최소화 하는 모델을 찾기 위해 다음과 같이 제안한다. 먼저, 비정합의 가능성이 높은 정합 쌍을 찾아내면 복잡도가 높은 비정형 정합 알고리즘을 수행하지 않아도 된다. 이를 위해 비정합의 가능성이 높은 경우를 찾기 위해 훈련 셋에서 분석한 결과, 올바른 정합 쌍의 개수가 4보다 작은 경우에는 비정형 정합을 수행하여도 올바른 결과를 가져 오지 못함을 발견하였다. 따라서 올바른 정합 쌍의 개수가 4 보다 작은 경우는 비정합이라고 판단하여 비정형 정합에서 우선적으로 제외한다. 올바른 정합 쌍의 개수가 4 보다 많을 경우에는 두 모델의 평균의 차이가 가장 큰 올바른 정합 쌍의 개수와 분산의 차이가 가장 작은 특징 정합 쌍의 가중치와의 비율을 식 9와 같이  $r$  을 계산한다.

$$r = \frac{\text{inliers}_n}{\text{weight}}. \quad (9)$$

훈련 영상에서 영상 정합 쌍에 대해서  $r$  을 계산하고 이를 정규 분포 모델로 구성하여 베이시안 분류를 통해 최소 오류를 갖는  $t$  값을 구한다. 그림 5는  $\text{inlier}_n > 4$  보다 큰 정합 쌍에 대해서 비정합( $\omega_1$ )과 비정형( $\omega_2$ )의 정규 분포를 도식하였다. 이를 이용하여 최소 오류를 갖는  $t$  을 찾고,  $t$  보다 큰 값을 갖는 쌍은 비정형일 가능성이 높다고 판단하여 비정형 정합 방법을 수행한다.

## 5. 비정형 객체 영상 정합 방법

비정형 객체 영상을 정합하는 다양한 방법 중에 특징 점을 사용하여 정합 쌍들의 유사성을 비교하고 이들을

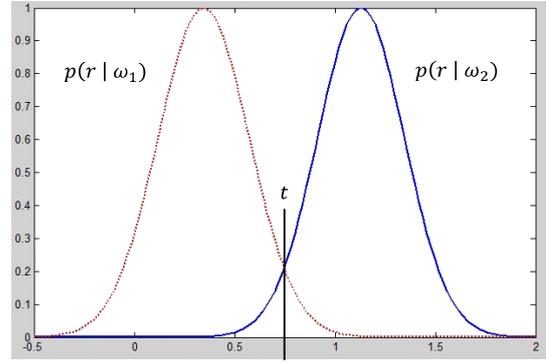


그림 5. 비정합과 비정형 정합 쌍의 정규 분포

Fig. 5. Gaussian distribution model of non-rigid and non-matching-pair

군집하는 Agglomerative Correspondence Clustering (ACC)<sup>[8]</sup>이 있다. 그러나 이 방법은 실시간으로 군집을 수행하여 초기 정합 쌍이 많을수록 복잡도가 높은 단점이 있다. 따라서 본 논문은 유효한 특징들을 선택하는 알고리즘을 적용하여 초기 정합 쌍의 개수를 줄여 속도를 개선한 Improved ACC<sup>[9]</sup>을 사용하였다.

특정 정합 쌍들의 연관성을 계산하기 위해 정합 쌍들의 기하학적 차이점(dissimilarity)을 계산한다. 1:N 정합 쌍에서 두 개의 정합 쌍을  $m_i = (x_i, x'_i, H_i)$  와  $m_j = (x_j, x'_j, H_j)$  라고 했을 때, 특징점의 서술자들 간의 차이를 유클리디안 거리로 나타낸 것을 광학적 차이점(Photometric dissimilarity)  $d_{app}(m_i)$ 로 나타내고, 정합 쌍을 구성하는 두 개의 특징점 간의 기하학적 차이점(Geometric dissimilarity)  $d_{gm}(m_i, m_j)$ 로 나타냈을 때, 두 합을 정합 쌍의 차이점으로 다음 식 10와 같이 정의한다.

$$\begin{aligned} d(m_i, m_j) &= d_{gm}(m_i, m_j) + \alpha \max(d_{app}(m_i, m_j)), \\ d_{gm}(m_i, m_j) &= \frac{1}{2}(d_{gm}(m_j|m_i) + d_{gm}(m_i|m_j)), \\ d_{gm}(m_j|m_i) &= \frac{1}{2}(|x'_j - H_i x_j| + |x_j - H_i^{-1} x'_j|), \\ d_{gm}(m_i|m_j) &= \frac{1}{2}(|x'_i - H_j x_i| + |x_i - H_j^{-1} x'_i|). \end{aligned} \quad (10)$$

각 정합 쌍들에 대한 군집은 kNN 연결 모델을 변형한 Adaptive Partial 연결 모델을 사용한다. 두 개의 군집  $C_a$ 와  $C_b$ 가 있을 때,  $D(C_a, C_b)$ 를 Cluster Dissimilarity Function(CDF)으로 정의하고, CDF와 연결 알고리즘을 통해 각 정합 쌍들을 군집한다. Adaptive



그림 6. TPS 변형을 이용한 인위적 비정형 영상과 의류 영상의 예제

Fig. 6. Example of Synthetic deformable images using TPS transformation and Clothing images.

Partial 연결 모델 식을 식 11로 정의하였다.

$$D_{AP}(C_a, C_b) = D_{kNN}(k, C_a, C_b),$$

$$\begin{cases} k = k_{AP} & \text{if } |C_a||C_b| \leq k_{AP}/\gamma_{AP} \\ k = \gamma_{AP}|C_a||C_b| & \text{else} \end{cases}$$

$$D_{kNN}(k, C_a, C_b) = \frac{1}{|\Gamma|} \min_{(i,j) \in \Gamma} d(m_i, m_j),$$

$$s.t. \Gamma \subset C_a \times C_b, |\Gamma| = \min(k, |C_a||C_b|). \quad (11)$$

$k_{AP}$ 와  $\gamma_{AP}$ 는 조정 파라미터이다.  $|C_a||C_b|$ 가  $k_{AP}/\gamma_{AP}$  보다 작으면 AP 연결은 kNN 연결과 동일하게 동작하고, 크면  $\gamma_{AP}|C_a||C_b|$ 의 정합 쌍을 사용한다.

### III. 실험

본 논문에서 제안된 정형과 비정형 객체 영상의 정합 방법을 검증하기 위해 스탠포드 대학의 Stanford Mobile Visual Search(SMVS)<sup>[10]</sup> 데이터 셋과 비정형 영상인 의류 영상을 사용하였다. SMVS 영상은 MPEG-7 CDVS<sup>[13]</sup>의 표준 영상에 포함되어 있으며, CD/DVD/책 표지, 랜드마크, 건물, 프린트, 그림 등의 정형 객체 영상들을 포함하고 있으며, 비정형 객체 영상 정합 방법의 성능 평가를 위해 SMVS의 질의 영상들을 Thin-Plate-Spline(TPS) 변형 알고리즘을 사용하여 변형 정도를 구분하여 3단계로 사용하였다. 그림 6은 SMVS 영상들과 이를 TPS를 이용하여 비정형화 시킨 영상과 자연 영상에서 발생할 수 있는 비정형 객

체의 대표적인 의류 영상의 예시 영상이다. 실험은 SMVS 원본을 정형 정합, 강도별로 변형시킨 TPS를 비정형 정합, 그리고 의류 영상을 자연영상으로 정의하고 최종적으로 모든 영상들을 혼합한 데이터 셋을 대상으로 성능 평가를 수행하였다. 성능 평가를 위해 TPR, FPR, ACC를 기준으로 사용하였으며, 그 정의는 다음과 같다.<sup>[11]</sup>

- True Positive Rate (TPR)

$$TPR = \frac{TP}{TP+FN} = \frac{TP}{P}$$

- False Positive Rate (FPR)

$$FPR = \frac{FP}{FP+TN} = \frac{FP}{P}$$

- Accuracy(ACC)

$$ACC = \frac{TP+TN}{P+N}$$

가. 비정형/비정합 영상의 구분 성능 검증

정합을 수행하기에 앞서 제안된 통계적 분석을 이용한 비정합과 비정형 영상의 구분 방법에 대한 성능을 평가하였다. 훈련 영상으로 정형 UKY 데이터 셋<sup>[12]</sup>을 사용하였다.  $t$  값의 변화에 따른 구분 성능은 다음 표 1과 같다. 결과에서 볼 수 있듯이  $t$  값이 커질수록 비정합 영상의 정확도는 높아지고, 오류율은 낮아진다. 반대로  $t$  값이 작을수록 비정형 영상의 정합도는 높아지고, 오류율은 낮아진다. 결론적으로  $t$  값이 작으면 비정형 영상으로 구분될 가능성이 높아져 전체적인 정합 성능

표 1. 통계적 분석을 이용한 비정합과 비정형 분리 성능 결과

Table 1. Separation Performance results of Deformable and non-matching image using statistical analysis.

t 값	비정형 영상		비정합 영상	
	Correct	Error	Correct	Error
0.30	<b>89.00%</b>	<b>11.00%</b>	80.89%	19.11%
0.40	88.89%	11.11%	81.14%	18.86%
<b>0.50</b>	<b>87.50%</b>	<b>12.50%</b>	<b>83.03%</b>	<b>16.97%</b>
0.55	85.89%	14.11%	84.25%	15.75%
0.60	83.67%	16.33%	86.08%	13.93%
0.65	81.08%	18.92%	89.64%	12.44%
0.70	77.31%	22.69%	89.64%	10.36%
0.75	72.83%	27.17%	<b>91.42%</b>	<b>8.58%</b>

은 올라가겠지만, 비정합 영상을 비정형으로 판단할 오류는 높아져서 전체적인 복잡도는 올라가게 될 것이다. 최적의 성능을 갖는 t 값을 결정하기 위해 전체 결과의 평균 오류율이 가장 낮은 0.75 를 선정하여 실험을 수행하였다.

나. 정형/비정형 영상 정합 방법

제안된 정합 방법의 성능 평가를 위해 다양한 영상 정합 방법들과 비교하였다. 단, 정형과 비정형을 동시에 정합하는 방법은 현재로선 발표된 것은 없으므로, 정형과 비정형 영상을 정합하는 대표적인 방법들과 비교한

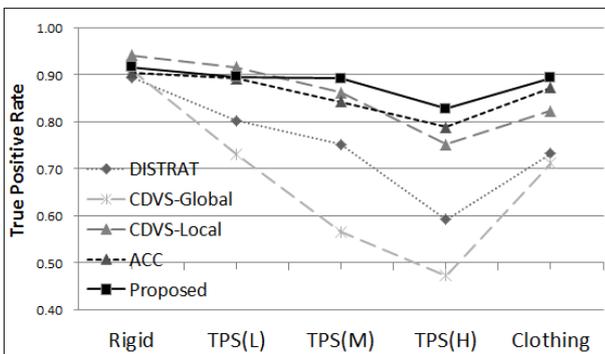


그림 7. 정합 방법과 제안된 방법과의 정합 성공률 비교 결과

Fig. 7. TPR comparison of proposed matching method and existing method.

다. 정형 영상 정합 방법으로 DISTRAT과 MPEG-7 CDVS을 사용하였다. CDVS는 2015년 국제표준화 채택을 앞둔 영상의 식별을 위한 정합과 검색 방법으로 전역(Global) 서술자와 지역(Local) 서술자로 구분된다. 또한, 비정형 객체 정합 방법으로는 특징점을 기반으로 정합을 수행하는 ACC와 비교하였다. 객관적인 성능 비교를 위해 FPR 을 5% 이내로 고정시키고, 정형 영상의 원본과 강도별 변형된 영상, 그리고 자연 영상인 의류 영상을 대상으로 정합 성공률을 비교하였다. 그림 7을 보면 DISTRAT과 CDVS의 경우 정형 영상에서는 좋은 성능을 보이지만 변형 강도가 커지면 정합 성공률은 급격히 낮아짐을 볼 수 있다. 또한 의류 영상의 경우는 비정형 정합 방법과 비교하여 15% 이상 차이가 나는 것이 확인된다. ACC의 경우 변형 강도에 따라 정합 성공률이 낮아지지만 차이가 적음을 알 수 있다. 제안된 방법은 변형 강도에 따른 정합 성공률의 차이가 가장 적음을 알 수 있고, 특히 강한 변형 강도와 자연 영상에서는 가장 좋은 성능을 보여준다.

그림 8은 전체 실험 영상에 대한 정확도를 나타낸다. 정확도는 정합 성공률과 실패율을 비교하여 종합적인 성능을 평가한다. 비정형 영상의 비율이 높으므로 정형 정합 방법은 대체로 정확도가 낮다. CDVS-Local과 ACC 그리고 제안된 방법이 거의 유사한 성능을 보여주지만 근소하게 제안된 방법이 높음을 확인하였다. 마지막으로 복잡도를 그림 9와 같이 비교하였다. 정합 성능에서 복잡도는 실시간에서 사용할 수 있는지를 판단해준다. 다만, 복잡도 평가에서 CDVS는 제외를 하였다. CDVS는 모바일 기반 정합 방법으로 압축된 서술자와 훈련에 의한 인덱싱 방법을 사용하므로 기타 방법과 비교하기엔 객관적이지 않다. 비정형 정합 방법인 ACC는 정합 판단시 정합 쌍들을 실시간으로 클러스터링을 하

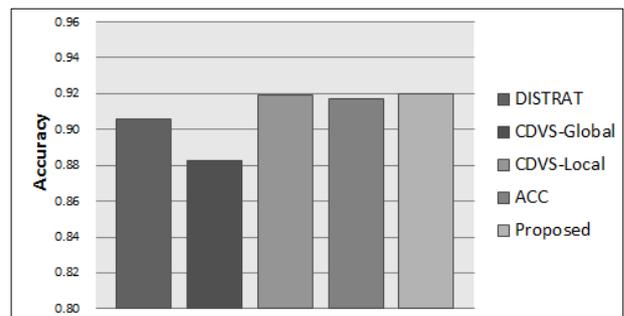


그림 8. 정합 방법에 따른 정확도 비교 결과

Fig. 8. Accuracy comparison of matching method.

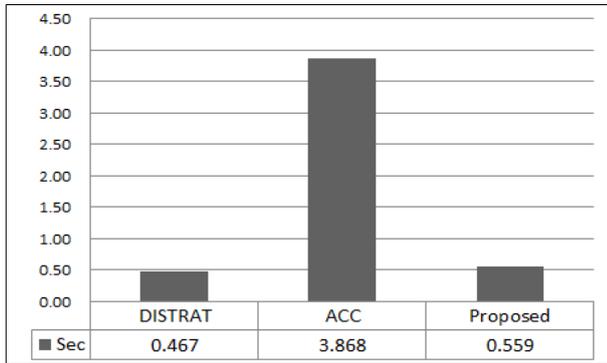


그림 9. 제안된 방법과 다른 정합 방법과의 복잡도 비교

Fig. 9. Complexity Comparison for proposed method with other matching methods

기 때문에 성능은 좋지만, 이로 인해 복잡도가 높다. 따라서 본 논문에서 제안하는 방법에 의해 비정합일 가능성이 높은 정합 쌍은 우선적으로 제외하는 방법을 사용하게 되면 불필요한 작업을 수행하지 않으므로, 전체 구조에서 복잡도가 줄어들게 된다. 결론적으로, 제안된 방법은 정형으로 판단되면 정형 정합으로 결정, 정형으로 판단되지 않을 때는 비정형일 후보군을 빠르게 결정하고 후보군에 대해서만 비정형 정합 방법을 수행하는 효과적인 구조를 갖는 것을 의미한다.

#### IV. 결 론

본 논문은 영상의 종류와 상관없이 동일한 특징을 사용하여 정형 객체 영상과 비정형 객체 영상을 정합할 수 있는 영상 정합 방법을 제안하였다. 기존의 정형 객체 영상 정합 방법과 비정형 객체 정합 방법은 목적이 다르므로 특징 추출부터 정합 방법까지 별도로 연구가 진행되어 왔다. 제안된 방법은 영상에서 동일한 특징을 추출하여 단일 구조로 정형과 비정형 객체 영상을 정합할 수 있게 된다. 영상 간 정합 쌍을 두 개로 구성하며 복잡도가 낮은 정형 정합을 우선적으로 수행하여 정합을 판단하고 정형 정합의 결과인 정합 정보를 통계적 분석을 통하여 비정형과 비정합의 경우를 효과적으로 분류할 수 있음을 입증하였으며 성능도 우수함을 확인하였다. 제안된 방법은 영상의 정합 정보를 이용하여 정형 정합의 가능성이 높은 경우와 비정합의 가능성이 높은 경우를 우선적으로 수행하거나 제외함으로써 전체 정합 방법에 효율을 가져왔다고 볼 수 있다. 정형 객체

영상과 비정형 객체 영상을 혼합한 실험 영상에서의 성능 평가 결과 정합 성능 면에서는 기존의 정형 정합과 비정형 정합 방법보다는 각각 4%와 1%의 근소하지만 좋은 성능을 나타내었으며, 복잡도 비교 결과 비정형 정합 방법과 비교하면 약 4배가 빨랐으며, 정형 정합 방법과는 거의 유사함을 나타내었다. 제안된 방법은 동일한 특징을 사용하여 정형과 비정형 영상 정합을 수행할 수 있으며, 성능은 유지하면서 복잡도가 낮은 장점이 있다. 향후 연구로 본 논문에서는 특징점 추출 방법으로 SIFT를 사용하였는데, SIFT는 정형 영상들에 최적화되어 있다는 단점이 있다. 따라서 비정형 영상들의 정합에서도 좋은 성능을 갖기 위한 특징을 연구하고, 비정형 객체 영상 데이터를 확보할 예정이다.

#### REFERENCES

- [1] CDVS, Call for Proposals for Compact Descriptors for Visual Search, ISO/IEC JTC1/SC29/WG11/N12201. Geneva, CH, March, 2011.
- [2] Lowe David G. "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 2 no. 60, pp 91-110, Nov. 2004.
- [3] J. Matas, O. Chum, M. Urban and T. Pajdla, "Robust Wide Baseline Stereo from Maximally Stable Extremal Regions," in Proc. 13th British Machine Vision Conference, pp. 384-393, 2002.
- [4] Mikolajczyk, Krystian, and Cordelia Schmid. "Scale & affine invariant interest point detectors." *International Journal of Computer Vision*, vol. 60, no. 1, pp 63-86, Oct. 2004.
- [5] Martin A. Fischler and Robert C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," in Proc. Image Understanding Workshop, pp. 71-88, 1980.
- [6] Lepsøy, G. Francini, G. Cordara, P. P. de Gusmao, "Statistical modelling of outliers for fast visual search," in Proc. International Conference on Multimedia and Expo, pp 1-6, Barcelona, Spain, July 2011.
- [7] R. J. Larsen and M. L. Marx, "An Introduction to Mathematical Statistics and its Applications," Prentice-Hall, 1986.
- [8] Cho, Minsu, and Jungmin Lee. "Feature

correspondence and deformable object matching via agglomerative correspondence clustering,” in Proc. International Conference on Computer Vision, pp 1280-1287, Kyoto, Japan, Sept. 2009.

[9] Hun-Jun Yang, In-Su Won, Dong-Seok Jeong, “On the Improvement of Deformable Object Matching,” Korea-Japan Joint Workshop on Frontiers of Computer Vision (FCV) 2014, pp 279-282, Okinawa, Japan, Feb. 2014.

[10] V. Chandrasekhar, D. M. Chen, S. S. Tsai, N. M. Cheung, H. Chen, G. Takacs, Y. Reznik, R. Vedantham, R. Grzeczuk, J. Bach, and B. Girod, “The Stanford mobile visual search dataset,” in Proc. second annual ACM conference on Multimedia systems, pp 117-122, San Jose, USA, Feb. 2011.

[11] Fawcett, T, “An introduction to ROC analysis,” *Pattern Recognition Letters*, vol 27, no. 8, pp 861-874, June 2006.

[12] D. Nister and H. Stewenius. “Scalable recognition with a vocabulary tree,” in Proc. of IEEE Conference on Computer Vision and Pattern Recognition, pp 2161-2168, New York, USA, June 2006.

[13] Test Model 11 : Compact Descriptors for Visual Search, ISO/IEC JTC1/SC29/WG11/N14222. San Jose, US, January, 2014.

— 저 자 소 개 —



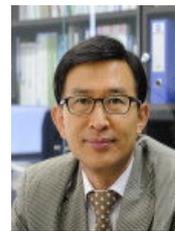
**원 인 수**(학생회원)  
 2006년 인하대학교 전자공학과 학사 졸업.  
 2009년 인하대학교 전자공학과 석사 졸업.  
 2009년~현재 인하대학교 전자공학과 박사 과정.  
 <주관심분야 : 영상신호처리, 컴퓨터비전, 영상 검색, 패턴인식>



**장 혁**(학생회원)  
 1999년 인하대학교 전자공학과 학사 졸업.  
 2000년 인하대학교 전자공학과 석사 졸업.  
 2013년~현재 인하대학교 전자공학과 박사 과정.  
 <주관심분야 : 통신, 컴퓨터, 신호처리, 반도체>



**양 훈 준**(학생회원)  
 2011년 인하대학교 전자공학과 학사 졸업.  
 2013년 인하대학교 전자공학과 석사 졸업.  
 2013년~현재 인하대학교 전자공학과 박사 과정.  
 <주관심분야 : 패턴인식, 비디오압축, 영상신호처리>



**정 등 석**(정회원)  
 1977년 서울대학교 전기공학과 학사 졸업.  
 1985년 Virginia Tech 전자공학과 공학 석사.  
 1988년 Virginia Tech 전자공학과 공학 박사.  
 1988년~현재 인하대학교 전자공학과 교수  
 1990년~1994년 전자공학회 논문지 편집위원  
 1990년~1994년 통신학회 논문지 편집위원  
 2000년~2004년 정보전자공동연구소 소장  
 2010년~2012년 인하대학교 IT공대학장  
 2012년 인하공업전문대학교 총장  
 <주관심분야 : 영상신호처리, 컴퓨터비전, 패턴인식, 내용기반 멀티미디어 검색>