논문 16-41-03-06 The Journal of Korean Institute of Communications and Information Sciences '16-03 Vol.41 No.03
http://dx.doi.org/10.7840/kics.2016.41.3.317

멀티 프레임 기반 건물 인식에 필요한 특징점 분류

박 시 영◆, 안 하 은**, 이 규 철**, 유 지 상°

Classification of Feature Points Required for Multi-Frame Based Building Recognition

Si-young Park◆, Ha-eun An**, Gyu-cheol Lee**, Ji-sang Yoo°

요 약

영상에서 의미 있는 특징점(feature point)의 추출은 제안하는 기법의 성능과 직결되는 문제이다. 특히 나무나 사람 등에서의 가려짐 영역(occlusion region), 하늘과 산 등 객체가 아닌 배경에서 추출되는 특징점들은 의미없는 특징점으로 분류되어 정합과 인식 기법의 성능을 저하시키는 원인이 된다. 본 논문에서는 한 장 이상의 멀티 프레임을 이용하여 건물 인식에 필요한 특징점을 분류하여 인식과 정합단계에서 기존의 일반적인 건물 인식 기법의 성능을 향상시키기 위한 새로운 기법을 제안한다. 먼저 SIFT(scale invariant feature transform)를 통해 일차적으로 특징점을 추출한 후 잘못 정합 된 특징점은 제거한다. 가려짐 영역에서의 특징점 분류를 위해서는 RANSAC(random sample consensus)을 적용한다. 분류된 특징점들은 정합 기법을 통해 구하였기 때문에 하나의 특징점은 여러 개의 디스크립터가 존재하고 따라서 이를 통합하는 과정도 제안한다. 실험을 통해 제안하는 기법의 성능이 우수하다는 것을 보였다.

Key Words : Occlusion region, multi-frame, feature extraction, feature matching, classification, homography, RANSAC

ABSTRACT

The extraction of significant feature points from a video is directly associated with the suggested method's function. In particular, the occlusion regions in trees or people, or feature points extracted from the background and not from objects such as the sky or mountains are insignificant and can become the cause of undermined matching or recognition function. This paper classifies the feature points required for building recognition by using multi-frames in order to improve the recognition function(algorithm). First, through SIFT(scale invariant feature transform), the primary feature points are extracted and the mismatching feature points are removed. To categorize the feature points in occlusion regions, RANSAC(random sample consensus) is applied. Since the classified feature points were acquired through the matching method, for one feature point there are multiple descriptors and therefore a process that compiles all of them is also suggested. Experiments have verified that the suggested method is competent in its algorithm.

※ 이 논문은 2016년도 정부(미래창조과학부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임 (No.R0132-16-1005, 온-오프라인에서의 콘텐츠 비주얼 브라우징 기술개발).
◆ First Author : Kwangwoon University Department of Department of Electronics, pksiyoung@kw.ac.kr, 학생회원
° Corresponding Author : Kwangwoon University Department of Department of Electronics, jsyoo@kw.ac.kr, 종신회원
* Kwangwoon University Department of Department of Electronics, mysco226@kw.ac.kr, 학생회원
** Kwangwoon University Department of Department of Electronics, lucifer_me@kw.ac.kr, 학생회원
논문번호 : KICS2015-10-321, Received October 1, 2015; Revised March 17, 2016; Accepted March 23, 2016

Ⅰ. 서 론

객체 인식 분야는 다양한 비전 응용에 적용 될 수 있는 기술이다. 한 예로 자동차에서 주변 영상을 입력 받아 사람을 인식하여 사고 발생률을 줄일 수도 있다. 특히 최근 모바일 폰의 활용 폭이 넓어지면서 내장된 카메라로 주변 물체들을 촬영하여, 획득된 영상내의 물체를 구별하는 다양한 프로그램들이 개발되고 있다. 이중에는 획득된 영상 내의 건물을 인식하고 인식된 건물을 자동으로 웹에서 검색하여 건물 내 매장의 다양한 정보를 사용자에게 제공하는 등의 서비스를 제공할 목적으로 특히 건물 인식과 관련된 연구가 활발히 진행되고 있다. 건물 인식을 위하여 입력으로 활용되는 영상의 형태는 크게 공중 영상(aerial image)과 도시 건물 영상(city building image)으로 구분 된다[1]. 도시 건물 영상은 스마트 폰에 내장된 카메라 등을 이용하여 획득된 영상으로 사람, 나무, 하늘 등의 주변 물체와 공존하는 경우가 대부분이고 랜드 마크와 같이 특이한 형태로 구성된 건물을 제외하면 대부분의 건물들이 비슷한 형태로 구성되어 있다. 따라서 건물 인식을 위해서는 먼저 특정 건물과 주변 환경의 분류 과정이 필요하며 이때 많은 데이터를 빠르게 처리하기 위한 기법이 필요하다. 또한 단순히 영상 정보만을 이용하지 않고 GPS 등의 정보를 이용하여 정보의 정확도를 높이는 방향으로 연구가 진행되고 있다[1].

건물 인식 시스템은 특징점 추출, 특징점 정합과 특징점의 분류의 총 세 단계로 나눌 수 있다. 특징점 추출 방식으로는 SIFT(scale invariant feature transform)[2]와 같이 글로벌 특징점(global feature)을 이용하는 방법과 건물의 기하학적 특징 또는 건물의 색상, 질감, 형상을 이용하는 내용기반 영상 검색(content based image retrieval) 방법이 있다[3]. 글로벌 특징점을 이용하여 정합을 할 경우 일반적으로 128차원이나 64차원의 디스크립터(descriptor)를 이용하는데 차원이 높은 정보를 사용할 경우 처리 시간이 길어진다. 이 문제를 해결하고자 PCA(principal component analysis)[4], LDA(linear discriminant analysis)[5], LPP(linear preserving projections)[6], SLPP(supervised LPP)[7], SDA(semi-supervised discriminant analysis)[8] 등의 기법에서는 차원을 낮추는 방법을 제시하고 있다.

특징점 기반 정합의 경우는 특징점이 많이 추출되지 않는다는 단점이 있기 때문에 최근에 내용기반 검색을 이용하는 정합 방식의 연구가 진행 중이다. 내용기반 검색은 문서위주의 텍스트 기반 자료 검색 방식에서 텍스트 대신에 영상의 색상, 질감, 형상과 같은 값들을 이용하여 유사한 영상을 찾는 방식이다[9]. 건물의 경우는 직선 형태가 많이 존재함으로 창문의 나열된 형태, 직선의 구조 등을 이용하여 건물의 특징점을 추출한다[10]. 특징점 정합은 참조영상(reference image)과 질의영상(query image)에서 추출한 특징점들 간의 거리(euclidean or mahalanobis distance) 차이를 가지고 상관관계를 찾아 정합한다.

마지막으로 특징점을 기반으로 건물을 분류하는 방법으로는 모든 건물들에서 추출한 특징점을 모아 SVM(support vector machine)[11]을 통해 학습 과정을 진행하고, 가설 함수를 설정하여 입력된 영상에서 건물을 인식하는 방법이 일반적이다. 위에서 설명한 건물 인식 방법은 주로 단일 건물의 인식에 대한 것이고, [12]에서는 다중 건물 인식 방법에 대해 소개하고 있다.

대부분의 건물 인식 기법들은 단일 영상에서 특징점들을 추출하고 이 특징점들을 디스크립터와 같이 고차원 정보로 변환하여 저장하게 된다. 특히 SIFT[2]의 경우 추출된 특징점은 건물 정보뿐만 아니라 중요하지 않은 배경 정보도 포함하기 때문에 특징점 정합의 정확도가 현저히 떨어진다. 본 논문에서는 인식하려는 건물에 대해 시점이 다르게 획득된 영상(멀티 프레임)들을 가지고 SIFT[2]에 적용하여 특징점들을 분류하고, 특징점 정합 과정의 정확도를 높이는 방법을 제안한다. 기존 방식에서는 하나의 영상을 입력으로 이용하고 이 입력 영상은 한 시점에서의 전체 건물의 형태를 포함하는 반면 제안하는 방식에서는 다른 각도에서 획득된 여러 view 영상을 입력으로 이용하기 때문에 다른 각도에서 바라본 건물의 모습이 각 입력 영상에 포함되게 된다. 입력된 영상들은 시점이 서로 달라 건물의 특징점은 여러 영상에서 나타나지만, 배경의 특징점은 한정된 영상에서만 나타난다. 따라서 RANSAC을 이용하여 특징점들의 연관 관계를 찾아 반복적으로 나타나는 특징점을 찾고 분류하는 방식을 통해 특징점 정합의 정확도를 높이고자 한다.

본 논문의 구성은 다음과 같다. 2장에서는 제안하고자 하는 건물 인식에 필요한 특징점 추출 과정과 분류 과정에 대해 설명한다. 3장에서는 실험을 통하여 제안하는 기법의 성능을 평가하며, 4장에서 결론을 맺는다.

Ⅱ. 본 론

본 논문에서 제안하는 멀티 프레임을 이용한 특징

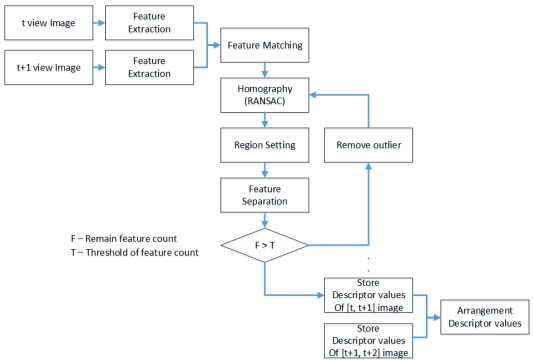


그림 1. 제안하는 멀티 프레임을 이용한 특징점 분류 기법의 흐름도
Fig. 1. Flowchart of the proposed feature classification algorithm with multi-frames

점 분류 기법의 흐름도는 그림 1과 같다. 특징점 분류 방식은 세 단계로 구성된다. 첫 번째 단계에서는 SIFT(scale invariant feature transform)[2] 기법을 적용하여 특징점을 추출한다. 두 번째 단계에서는 다수의 프레임을 이용하여 추출한 특징점들을 정합한다. 마지막 단계에서는 특징점 정합으로 획득한 다수의 특징점 쌍들을 RANSAC(random sample consensus)[13] 기법을 이용하여 호모그래피(homography) 행렬을 획득한 후, 이를 이용하여 건물에 필요한 특징점으로 분류하게 된다.

2.1 특징점 추출 및 정합

단일 영상에 대하여 특징점을 추출하는 방법은 SIFT(scale invariant feature transform)의 DOG(difference of gaussian)[2], SURF(speeded up robust feature)의 Haar wavelet[14], SUSAN(smallest uni-value segment assimilating nucleus test)[15]과 FAST(Features from Accelerated Segment Test)[16] 등 다양한 기법이 있다. 이 중 SIFT의 DOG[17]는 주변의 밝기 차이를 이용하며 기존에 존재하는 여러 기법들 중에서 연산량이 비교적 많지만 영상의 크기, 회전 변화와 잡음에 강인하고 반복성(repeatability)이 높다는 장점이 있다. 따라서 본 논문에서는 SIFT 기법을 이용하여 특징점 추출 및 정합을 수행한다. 먼저 특징점을 추출하기 위해 영상의 전체 영역에 DOG(difference of Gaussian) 기법을 적용한다[2]. 시점이 다른 영상으로부터 추출된 특징점들을 각 영상 간에 정합하기 위해서는 각각의 특징점에 대해서 기준이 되는 값을 먼저 정의할 필요가 있다. 4x4 크기의 블록(16화소)을 정의하고 하나의 특징점을 중심으로 총 16개의 주변 블록을 그림 2와 같이 설정한다[2].

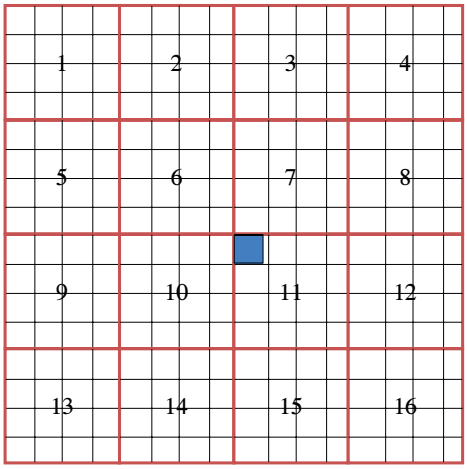


그림 2. SIFT의 디스크립터 블록 형태
Fig. 2. Blocks of SIFT for descriptors

이때 특징점은 그림 2의 파란색 부분처럼 11번째 블록의 처음 화소에 위치하게 한다. 4x4 크기의 16개 화소로 이루어진 각 블록에서 각 화소 값에 대한 수평 방향과 수직 방향의 일차 미분(인접화소 값의 차이) 값을 구하고, 수평, 수직 방향의 미분 값을 이용하여 각 화소 값의 기울기 방향과 기울기 크기를 구하게 된다. 이때 그림 3과 같이 각 화소의 기울기 방향은 0도부터 45도 간격으로 총 8개의 방향으로 양자화 하게 되고, 기울기 값도 원래 기울기의 크기에 따라 양자화 과정에서 가중치를 가지고 변환된다. 식 (1)은 화소의 기울기 방향이 0도와 45도사이의 값일 경우 양자화 과정에서 화소의 기울기의 크기가 0도와 45도 방향 성분으로 분리되는 과정을 보여주는 식이다[2].

$$
\begin{aligned}
&0 < x_{\angle} < 45 \\
&H_0 = \frac{x_{\angle} - 0}{45} x_{magnitude} \\
&H_{45} = \frac{45 - x_{\angle}}{45} x_{magnitude}
\end{aligned}
\tag{1}
$$

여기서 $x_{\angle}$은 화소의 기울기의 방향을 나타내고, H_0는 화소의 기울기의 크기 중 0도 방향의 크기를 나타내고, H_{45}는 화소의 기울기의 크기 중 45도 방향의 크기를 나타낸다.

결국 그림 3과 같이 각 블록 내의 16개 화소가 갖는 8개의 방향에 대한 기울기와 기울기 값을 더하여 블록별로 8개의 기울기 정보를 구할 수 있다. 따라서 하나의 특징점은 주변 16개 블록에 대해 각각 8개 방향의 기울기 정보를 가지게 됨으로 모두 128(8x16)차

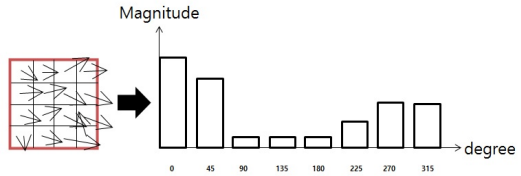


그림 3. 기울기 방향과 크기의 히스토그램 변환
Fig. 3. Conversion of the inclination directions and magnitudes into histogram

원의 디스크립터(descriptor)로 정의된다. 각 특징점의 디스크립터 정보(기울기 정보)간 유클리디언(euclidean) 거리 차이의 합이 가장 작은 특징점들끼리 정합이 된다. 그림 4는 SIFT[2]를 이용한 특징점 정합 결과를 보여준다. 정합이 잘 된 특징점들이 있는 반면에 정합이 잘못된 특징점들도 소수 존재하는 것을 확인할 수 있다. 그림 4에서 파란색 선으로 정합된 특징점은 정확하게 정합된 특징점이고, 빨간색 선으로 정합된 특징점은 잘못 정합된 특징점을 보여주고 있다.

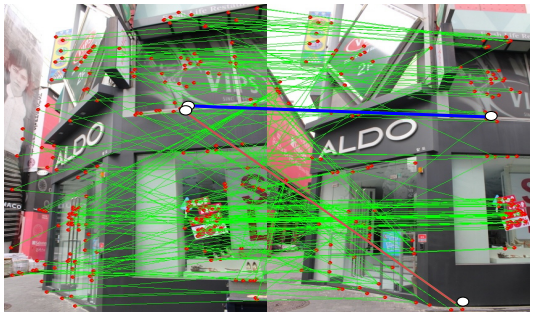


그림 4. SIFT 기법을 이용한 특징점 정합
Fig. 4. Feature point matching by the SIFT algorithm

2.2 호모그래피를 이용한 특징점 분류

정합이 잘못된 특징점들을 제거하기 위해서 호모그래피(homography) 기법을 이용하여 특징점을 분류한다. 호모그래피 기법으로 두 영상간의 기하학적인 관계를 알게 되면 특징점들을 정확하게 정합할 수 있으며, SIFT를 이용한 특징점 정합 과정에서 잘못 정합된 특징점들을 제거할 수 있다.

먼저 시점이 다른 두 영상에서 정합된 특징점 네 쌍을 선택한 다음 선택한 특징점들을 이용하여 호모그래피 행렬 H를 획득한다[18]. 이 경우 정합이 잘못된 특징점은 호모그래피 행렬의 값이 원래의 호모그래피 행렬과 전혀 다른 값을 갖게 된다. 잘못 정합된 특징점의 비율이 50% 미만인 경우 RANSAC[13] 기법을 이용하여 정확한 호모그래피 행렬을 획득할 수 있다. 무작위로 정합된 특징점 네 쌍을 선택하여 호모그래피 행렬 H를 획득한다[18]. 한 영상에서 획득된 각 특징점들과 행렬 H를 곱하면 대응되는 다른 영상의 특징점을 결정할 수 있다. 호모그래피를 이용하여 획득한 특징점과 SIFT를 이용하여 획득한 특징점간의 유클리디언 거리차이를 구하여 모두 합산한다. 이 과정을 다른 특징점들에 대해 반복하여 거리 차이의 합산 값이 최소가 되는 호모그래피 행렬 H를 구하면 시점이 다른 두 영상의 최종 호모그래피 행렬로 정의한다.

RANSAC 기법을 통해 구한 호모그래피 행렬 H를 가지고 정확하게 정합된 특징점을 분류하기 위해서 한 영상에서 정합된 특징점들을 식 (2)의 호모그래피 행렬 변환을 통해 변환된 좌표 값을 구한다.

$$\begin{bmatrix} X_2 \\ Y_2 \\ 1 \end{bmatrix} = H \begin{bmatrix} X_1 \\ Y_1 \\ 1 \end{bmatrix} \tag{2}$$

여기서 $[X_1, Y_1, 1]$은 한 영상의 좌표이고, $[X_2, Y_2, 1]$은 행렬 H에 의해 변환된 다른 영상의 정합된 특징점의 좌표이다. 변환된 좌표와 다른 영상에서 정합된 좌표와의 유클리디언 거리 차이를 식 (3)을 통해 계산한다.

$$|X_3 - X_2| < t, \; |Y_3 - Y_2| < t \tag{3}$$

여기서 X_3과 Y_3은 SIFT 기법을 통해 정합된 다른 영상의 특징점 좌표를 나타내고, t는 임의의 임계값을 나타낸다. 임계값이 작을수록 두 좌표의 거리 차이가 작기 때문에 정확하게 정합된 특징점으로 분류할 수 있다. 그림 5는 그림 4에서 획득한 정합된 특징점들을 이용하여 호모그래피 행렬 H를 계산한 후, 식 (3)을 통해 분류한 특징점들을(빨간 점, 초록색 선) 보여주고 있다. 이 경우의 특징점은 두 영상에서 잘 정

그림 5. 호모그래피 행렬을 이용한 특징점 분류
Fig. 5. Classification of features by the homography matrix

합된 특징점으로 생각할 수 있다.

분류된 특징점들은 그림 5에서 보는 바와 같이 육안으로도 정확하게 정합되어 있음을 확인 할 수 있다. 하지만 그림5에 보인 특징점들은 정확하게 정합된 특징점을 모두 포함하고 있지 않다. 이런 현상은 하나의 호모그래피 행렬이 영상 내의 같은 평면에 존재하는 특징점의 정합점만을 찾아낼 수 있다는 특성에 기인한 것이다.

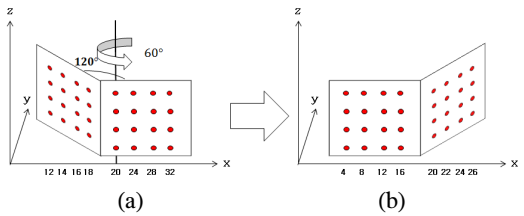
그림 6(a)는 3차원 공간에서 서로 평행하지 않은 두 평면을 120도의 차이를 두고, 거리간격이 4인 점들을 평면상에 위치시켰을 때 2차원 공간 측면에서 x축 값을 보여주고 있다. 그림 6(b)의 경우 x 좌표 20을 중심으로 60도로 z-축 회전을 시킬 경우 평면상에 위치한 점들이 2차원 공간 측면에서 변화된 값을 보여주고 있다. 그림과 같이 회전 시킬 경우 두 평면에 위치하는 점들은 2차원 공간 측면에서 x 좌표의 값이 서로 다른 것을 확인할 수 있다.

표 1은 그림 6(a)에 존재하는 두 평면을 x 좌표 값 20에서 z-축 방향으로 60도 회전할 때 변화되는 x 좌표의 값을 보여주고 있다. 표 1의 첫 번째 행은 그림 6(a)에 존재하는 빨간 점들의 x 좌표 값을 나타내고, 두 번째 행은 그림 6(b)에 보인 회전 후의 대응되는 점들의 x 좌표 값을 나타낸다. 세 번째 행은 회전으로 인해 변화된 x 좌표 값의 차이로 식 (4)를 이용해 구할 수 있고, 네 번째 행은 2차 변화량으로 정의하며 식 (5)을 이용하여 구할 수 있다.

$$X_{d_n} = X_{A_n} - X_{B_n} \tag{4}$$

$$X_{dd_n} = X_{d_{n+1}} - X_{d_n} \tag{5}$$

여기서 X_{B_n}과 X_{A_n}는 각각 회전전의 n 번째 x 좌표의 값과, 회전 후의 n 번째 x 좌표 값을 각각 의미한다. X_{d_n}는 n 번째 점에 대한 x 좌표의 일차 변화량을



(a) (b)

그림 6. z축을 중심으로 두 평면의 회전 (a) 회전 전 (b) 회전 후
Fig. 6. Rotation of the planes about the z-axis (a) Before rotation (b) After rotation

표 1. 회전에 따른 x 좌표 값의 변화량
Table 1. Variation of x coordinate value after rotation

Before rotation	12	14	16	18	20	24	28	32
After rotation	4	8	12	16	20	22	24	26
1st variation	-8	-6	-4	-2	0	2	4	6
2nd variation	-	-2	-2	-2	-2	2	2	2

나타내고, X_{dd_n}는 n 번째 이차 변화량으로 정의된다.

표 1에 따르면 서로 평행하지 않은 평면들을 3차원 공간에서 회전시키고 2차원 공간에서 점들의 좌표 값의 상관관계를 분석하였을 경우, 평면들을 나누는 기준이 되는 x축 값 20을 중심으로 두 평면들이 서로 다른 이차 변화량을 가지고 있음을 확인 할 수 있다. 표 1의 분석 결과를 이용하여 특징점들의 이차 변화량을 계산하고 평행하지 않는 평면의 존재 여부를 확인할 수 있다.

본 논문에서는 SIFT 기법과 RANSAC 기법을 이용하여 구한 하나의 호모그래피 행렬 H를 적용하여 표 1과 같이 한 영상의 좌표들을 변환시켜 이차 변화량이 갑자기 변하는 구간이 있는지 확인한다. 식 (6)과 (7)은 한 영상의 점들을 호모그래피 행렬 H를 통해 변환된 좌표를 보여준다.

$$H = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}, \quad \begin{bmatrix} X_2 \\ Y_2 \\ 1 \end{bmatrix} = H \begin{bmatrix} X_1 \\ Y_1 \\ 1 \end{bmatrix} \tag{6}$$

$$X_2 = \frac{aX_1 + bY_1 + c}{gX_1 + hY_1 + i}, \; Y_2 = \frac{dX_1 + eY_1 + f}{gX_1 + hY_1 + i} \tag{7}$$

여기서 H는 RANSAC을 통해 구한 초기 호모그래피 행렬 H로 임의의 원소 $[a, b, c, d, e, f, g, h, i]$로 구성된다. $[X_1, Y_1, 1]$은 한 영상의 좌표이고, $[X_2, Y_2, 1]$은 행렬 H에 의해 변환된 좌표이다. 식 (8)~(15)는 식 (6)의 행렬 H의 원소 값을 가지고 x축에 대한 이차 변화량을 구하는 과정을 보여주고 있다.

$$\begin{aligned} X_{d_n} &= X_{2_n} - X_{1_n} \\ &= \frac{aX_{1_n} + bY_{1_n} + c}{gX_{1_n} + hY_{1_n} + i} - X_{1_n} \\ X_{d_{n+1}} &= X_{2_{n+1}} - X_{1_{n+1}} \end{aligned} \tag{8}$$

$$= \frac{aX_{1_{n+1}} + bY_{1_{n+1}} + c}{gX_{1_{n+1}} + hY_{1_{n+1}} + i} - X_{1_{n+1}} \quad (9)$$

$$X_{1_n} = X_{1_n} + 1,\ Y_{1_n} = Y_{1_n} \quad (10)$$
$$X_{d_{n+1}} = X_{2_{n+1}} - X_{1_{n+1}}$$

$$= \frac{a(X_{1_n}+1) + bY_{1_n} + c}{g(X_{1_n}+1) + hY_{1_n} + i} - (X_{1_n}+1) \quad (11)$$

$$X_{d_{n+1}} - X_{d_n} = \frac{a(X_{1_n}+1) + bY_{1_n} + c}{g(X_{1_n}+1) + hY_{1_n} + i} - \frac{aX_{1_n} + bY_{1_n} + c}{gX_{1_n} + hY_{1_n} + i} - 1 \quad (12)$$

$$X_{d_{n+2}} - X_{d_{n+1}} = \frac{a(X_{1_n}+2) + bY_{1_n} + c}{g(X_{1_n}+2) + hY_{1_n} + i} - \frac{a(X_{1_n}+1) + bY_{1_n} + c}{g(X_{1_1}+1) + hY_{1_1} + i} - 1 \quad (13)$$

$$X_{dd_n} = \frac{a(X_{1_n}+n+1) + bY_{1_n} + c}{g(X_{1_n}+n+1) + hY_{1_n} + i} - \frac{a(X_{1_n}+n-1) + bY_{1_n} + c}{g(X_{1_n}+n-1) + hY_{1_n} + i} \quad (14)$$

$$X_{dd_n} = \frac{a(n+1)+c}{g(n+1)+i} - \frac{a(n-1)+c}{g(n-1)+i} \quad (X_{1_1} = 0,\ Y_{1_1} = 0) \quad (15)$$

여기서 X_{2_n}은 n 번째 x축의 좌표 값, X_{d_n}은 X_{2_n}의 일차 변화량을 나타내고 X_{dd_n}은 X_{2_n}의 이차 변화량을 나타낸다. 초기 좌표를 $(0,0)$으로 가정하면 X_{1_1}, Y_{1_1}은 각각 0이 되고, z-축으로 회전을 한다고 가정하면 x축의 대한 이차 변화량은 최종적으로 식 (15)과 같이 표현된다. 식 (15)에서 행렬 H의 원소들의 값이 일정하므로, 이차 변화량의 값은 점점 증가하거나 감소하는 경우만 존재하게 된다.

이 방법을 이용하면 영상에 존재하는 평행하지 않는 다수의 평면에 대해 각각의 평면에 대응되는 호모그래피 행렬을 구하여 정확히 정합된 특징점을 분류할 수 있다. 2.3절에서 이 기법에 대한 설명을 하고 있다.

2.3 다른 평면상의 특징점 분류 및 특징점 제거

다른 평면이 존재할 경우 대응하는 호모그래피 행렬을 구하는 방법은 2.2절에서 제안한 방법과 비슷하다. 2.2절에서 한 평면의 호모그래피 행렬으로 정확하게 정합된 특징점을 제외한 나머지 특징점을 가지고 RANSAC[13]기법을 이용하여 다른 평면에 대한 호모그래피 행렬을 구한다. 이 경우 고려해야 할 사항이 있다. 앞에서도 언급했듯이 RANSAC은 잘못 정합된 특징점의 비율이 50% 미만일 때 효과적으로 호모그래피 행렬을 구할 수 있다. 기존의 SIFT 기법을 통해 정합된 특징점은 정확히 정합된 특징점과 잘못 정합된 특징점이 모두 존재 하는데, 일차적으로 정확하게 정합된 특징점이 분류되기 때문에 남은 특징점들 중 잘못 정합된 특징점의 비율은 증가하게 된다. 따라서 다른 평면상의 특징점을 분류하기에 앞서 잘못 정합된 특징점을 제거하여 그 비율이 50% 미만으로 유지되도록 하는 과정이 필요하다.

먼저 앞에서 정확하게 정합된 특징점들이 포함된 영역을 설정한다. 그림 7은 그림5의 잘 정합된 특징점들이 포함된 영역(검은색 사각형)을 보여주고 있다. 그림 7에서 설정한 영역은 분류된 특징점들 중에서 x축과 y축의 각각 최소 값과 최대 값을 찾아 사각형 형태로 설정해준 영역이다.

설정한 영역의 크기를 영역이 포함된 평행한 하나의 평면과 비슷하게 설정해 주기 위해서 초기 호모그래피 행렬에 의해 정확하게 정합된 특징점을 이용한다. 이 특징점들은 같은 평면(평행한 평면)에 포함되어 있기 때문에 이 점들을 가지고 기존의 호모그래피 행렬보다 정확한 호모그래피 행렬을 구할 수 있다. 이 행렬을 바탕으로 식 (2)와 식 (3)을 가지고 정확하게 정합된 특징점을 다시 분류하여 영역을 재설정한다. 그림 8(a)는 초기에 설정한 영역을 보여주고 8(b)는

그림 7. 특징점이 포함된 영역 설정
Fig. 7. Decision of the region with feature points

(a) (b)

그림 8. (a) 초기 설정한 영역 (b) 재설정한 영역
Fig. 8. (a) Initial region (b) Expanded region

재설정한 영역(검은색 영역)을 보여주고 있다.

영역을 재설정한 이후 잘못 정합된 특징점을 찾게 된다. 그림 8(a)처럼 초기에 설정한 영역보다 그림 8(b)처럼 확장된 영역에서 잘못 정합된 특징점을 상대적으로 많이 찾을 수 있다. 그림 9는 영역을 재설정한 이후 제거된 특징점을 보여주는 그림으로 초록색 선은 정확히 정합된 특징점을 보여주고 빨간 선은 잘못 정합된 특징점을 나타낸다.

호모그래피 행렬을 통해서 처음에 분류된 정합 특징점들과 영역 설정을 이용해서 찾은 잘못 정합된 특징점들을 제외한 특징점들을 가지고 다시 RANSAC 기법을 이용하여 다른 평면에 존재하는 정합된 특징점을 찾는 과정을 반복하게 된다. 그림 10는 이 방법으로 추가적으로 찾은 정합된 특징점들을 보여 주고 있다.

앞의 방법을 충분히 반복함으로써 영상에서 존재하는 여러 평면에 대해서 정확히 정합되는 특징점을 분류할 수 있다. 반복하는 횟수가 증가 할수록 남아있는 특징점의 개수가 작아지면 RANSAC의 기본 조건(잘못 정합된 특징점의 비율)의 신뢰도가 떨어진다. 따라서 남아있는 특징점의 개수가 N개 이하로 내려가는 경우 반복 과정을 중단한다. 그림 11(a)는 전체 정합된 특징점들을 보여주고 11(b)는 제안한 방법을 통해

그림 9. 잘못 정합된 특징점 제거
Fig. 9. Elimination of the mismatching feature points

그림 10. 재 분류된 특징점
Fig. 10. Classification of the remaining feature points

(a) (b)

그림 11. (a) 전체 특징점 정합 쌍 (b) 제안한 방법을 통해 정확히 정합된 정합 쌍
Fig. 11. (a) Total matched pairs of the feature points (b) Well matched pairs of the feature points by the proposed algorithm

최종적으로 정확하게 정합된 것으로 분류된 특징점(빨간점, 초록선)들을 보여주고 있다.

2.4 분류된 특징점의 디스크립터 정보 정리

본 논문에서는 2.2절과 2.3절의 과정을 통해서 한 건물의 모든 참조 영상들에 대해 특징점 분류 과정을 진행한다. 분류된 하나의 특징점의 디스크립터 정보는 그 특징점이 존재하는 참조영상의 개수, 즉 최소 2개부터 모든 참조 영상의 수만큼 존재할 수 있다. 하나의 특징점이 존재하는 참조영상의 수만큼 128차원의 디스크립터가 존재하기 하는데 이 경우 디스크립터의 정보는 중복될 수도 있고 질의 영상 특징점과의 정합 과정 수행시간도 증가할 수 있다. 이 문제는 하나의 특징점이 존재하는 참조영상의 수만큼 존재하는 디스크립터 정보를 대응하는 하나의 디스크립터 정보로 정리함으로써 해결할 수 있다.

다음은 다수의 디스크립터 정보를 하나의 디스크립터 정보로 정리하는 5단계 과정이다.

Step 1: 특징점의 초기 디스크립터의 값은 해당 특

징점의 디스크립터 정보로 저장한다.

Step 2: 해당 특징점의 디스크립터가 저장 되어 있을 경우 저장된 디스크립터 정보에 따라 128차원의 각 차원의 임계값을 식 (16)에 의해 설정한다.

$$x_d^n/20+5=t_d^n \tag{16}$$

여기서 x_d은 디스크립터의 d번째 차원의 값을 나타내고, t_d는 d번째 차원의 임계값을 나타낸다.

Step 3: 다음으로 입력되는 디스크립터의 정보가 식 (17)에 만족할 경우 해당 차원의 값을 식 (18)에 의해 저장한다.

$$\left|x_{d_n}^n \pm t_{d_n}^n\right| \geq x_{d_{n+1}}^n \tag{17}$$

$$\frac{x_{d_k}^n \times n + x_{d_k}^{n+1}}{n+1} \tag{18}$$

여기서 $x_{d_k}^n$는 n번째 참조영상 특징점의 k번째 디스크립터의 값을 나타낸다.

Step 4: 식 (17)의 경우를 만족하지 않을 경우 현재의 디스크립터 정보와 입력으로 들어온 디스크립터의 정보를 모두 저장한다. 이미 두 가지 정보를 가지고 있을 경우 세 가지의 정보 중 차이가 작은 두 정보를 저장한다.

Step 5: 같은 특징점이 존재하는 모든 참조영상에 대해 128개의 디스크립터 정보 모두에 대해 step 1~4의 과정을 진행한 이후 최종적으로 두 가지 정보를 가지고 있는 차원은 주변의 차원과 비교하여 차이가 작은 값을 저장한다.

Ⅲ. 실 험

실험 영상은 ETRI에서 제공한 명동 DB를 사용한다. 실험 영상은 참조 영상과 질의 영상으로 구성되며, 영상의 해상도는 427x640과 640x427이다. 실험 환경은 Microsoft사의 Microsoft Visual Studio C++ 2010과 OpenCV 2.4.8 라이브러리를 이용하여 구현하고, 3.40GHz의 인텔 i5 쿼드코어 프로세서를 이용한다.

본 논문에서는 재안하는 기법의 정합 결과의 성능을 파악하기 위해 기존의 SIFT 기법과 recall 값을 비교한다. Recall 값이란 전체 정합 쌍에서 정확하게 정합된 비율을 나타내는 값으로 식 (19)으로 정의된다.

$$recall = \frac{N\ correct\ matches}{N\ correspondences} \times 100\% \tag{19}$$

여기서 N correspondences 는 최종적으로 분류된 정합 쌍의 전체 개수이고, N correct matches 는 육안으로 확인 했을 때 정확하게 정합 된 특징점의 개수이다.

그림 12는 A 건물의 참조 영상들로써 총 18장으로 이루어져 있고, 일정한 간격을 두고 시점이 다른 영상들로 구성되어 있다. 그림 13은 26장의 A 건물 질의 영상으로, 참조 영상과 다르게 조명 변화와 가려짐 영역(occlusion region) 등이 포함되어 있다.

표 2는 A 건물의 참조 영상들을 제안한 과정을 통해 특징점 분류를 수행한 시간을 보여준다. 표 2의 결과에 따르면 두 참조 영상을 가지고 특징점을 분류하는 과정은 약 2초의 수행 시간이 필요하다.

표 3은 A 건물의 참조 영상과 질의 영상에 대해 제

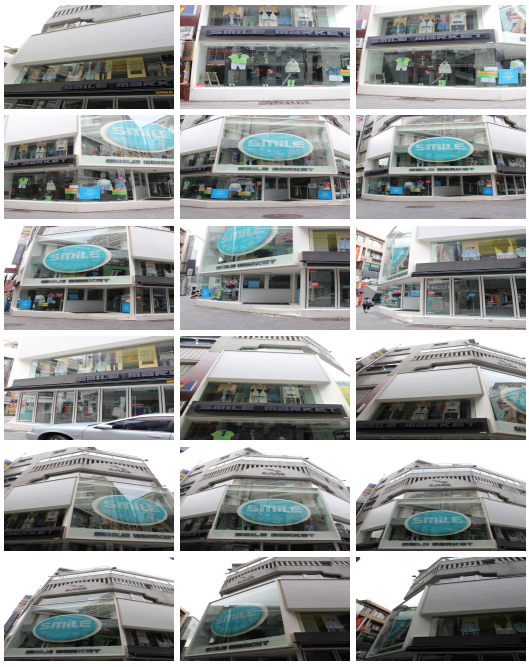


그림 12. 명동 A 건물 참조 영상
Fig. 12. Reference images of Myeong-dong A building

그림 13. 명동 A 건물 질의 영상
Fig. 13. Query images of Myeong-dong A building

표 2. A 건물 참조 영상의 수행 시간
Table 2. Processing time of A building's reference images

Building (Reference)	Processing time(s)	Building (Reference)	Processing time(s)
A(1, 2)	3.602	A(10, 11)	2.460
A(2, 3)	2.466	A(11, 12)	1.970
A(3, 4)	2.828	A(12, 13)	1.881
A(4, 5)	2.726	A(13, 14)	2.198
A(5, 6)	2.179	A(14, 15)	2.202
A(6, 7)	2.645	A(15, 16)	2.248
A(7, 8)	2.541	A(16, 17)	1.683
A(8, 9)	2.636	A(17, 18)	1.409
A(9, 10)	2.336		

안한 기법의 recall 값을 SIFT의 recall 값과 비교한 결과이다. SIFT를 적용한 경우의 recall 값은 하나의 질의 영상과 대응하는 특징점이 존재하는 경우 각각의 참조 영상 디스크립터 정보와의 정합 결과를 평균한 결과이다.

표 3. A 건물 특징점 정합 (recall in %)
Table 3. Feature Matching of A building (recall in %)

Building (Query)	Method	
	SIFT(%)	Proposed algorithm(%)
A(1)	60.78	80.24
A(2)	36.84	50.59
A(3)	12.67	30.25
A(4)	22.22	44.26
A(5)	22.22	43.28
A(6)	9.52	13.25
A(7)	9.43	18.25
A(8)	15.22	22.35
A(9)	1.51	15.54
A(10)	1.33	16.66
A(11)	17.14	48.52
A(12)	50.98	70.58
A(13)	0.00	3.58
A(14)	8.33	13.35
A(15)	34.09	54.25
A(16)	13.84	38.20
A(17)	25.92	37.34
A(18)	13.92	23.32
A(19)	13.15	31.33
A(20)	26.92	45.52
A(21)	31.75	44.75
A(22)	31.58	49.32
A(23)	22.67	50.25
A(24)	12.20	32.32
A(25)	18.27	31.00
A(26)	14.61	28.35

표 3의 결과를 보면, 표 3의 recall 값이 전체적으로 낮은 경우가 많은데 그 이유는 참조 영상과 질의 영상 간의 영상의 밝기 차이가 크기 때문이다. 두 가지 방법 모두 recall 값이 높게 측정된 질의 영상 A(1), A(12)번의 경우 시점이 건물의 정면으로 획득한 영상이고, 비슷한 시점의 참조 영상이 존재하는 경우이다. 질의 영상A(9), A(10), A(13), A(14) 번의 경우 SIFT의 recall 값과 제안한 방법의 recall 값 모두 작은 것을 확인 할 수 있다. 이 경우 영상을 획득한 시점이 옆에서 바라보며 획득한 영상이고, 밝기 차이가 심한 경우이다.

표 3의 결과를 보면 제안한 방법의 recall 값이 SIFT 기반의 특징점 추출을 통한 recall 값보다 전체적으로 10-30% 이상 높은 것을 확인할 수 있다.

그림 14. 분류된 특징점
Fig. 14. Classified feature points

Ⅳ. 결 론

본 논문에서는 영상 인식과정에서 멀티 프레임을 이용하여 특징점을 분류하고 정합의 성능을 향상시키는 새로운 기법을 제안하였다. 기존의 SIFT(scale invariant feature transform) 기반으로 건물의 특징점을 추출하는 기법은 인식하고자 하는 객체외의 주변 환경에 대해서도 특징점이 추출되기 때문에 정합의 정확성이 떨어지는 경우가 빈번하였다. 본 논문에서 제안하는 기법에서는 일정한 간격으로 시점이 다른 영상들을 이용하여 호모그래피(homography) 행렬을 획득하고 RANSAC(random sample consensus) 기법을 적용하여 특징점을 분류하였다. 하나의 호모그래피 행렬은 영상의 한 평면의 특징점만 분류하기 때문에 분류되고 남은 특징점을 가지고 호모그래피 행렬을 구하는 과정을 반복함으로써 영상에서 정확하게 정합된 특징점을 모두 분류한다. 제안하는 기법의 recall 값은 SIFT를 이용한 recall 값보다 향상시킬 수 있었다.

References

[1] J. Li, W. Huang, L. Shao, and N. Allinson, "Building recognition in urban environments: A survey of state-of-the-art and future challenges," *Inf. Sci.*, vol. 277, no. 1, pp. 406-420, Sept. 2014.

[2] D. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Computer Vision*, vol. 60, no. 2, pp. 91-110, Nov. 2004.

[3] Y. Li and L. G. Shapiro, "Consistent line clusters for building recognition in CBIR," in *Proc. 16th Int. Conf. Pattern Recognition,* vol. 3, pp. 952-956, Aug. 2002.

[4] I. T. Jolliffe, *Principal component analysis*, 2nd Ed., Springer, 2002.

[5] G. J. Malachlan, *Discriminant analysis and statistical pattern recognition*, Wiley-interscience, New York, 1992.

[6] X. He and P. Niyogi, "Locality preserving projection," in *Proc. Conf. Advances in Neural Inf. Process. Syst.*, 2003.

[7] D. Cai, X. He, and J. Han, *Using graph model for face analysis*, Department of Computer Science, University of Illinois at Urbana Champaign, Sept. 2005.

[8] D. Cai, X. he, and J. Han, "Semi-supervised discriminant analysis," in *Proc. IEEE 11th Int. Conf. Computer Vision,* pp. 1-7, Oct. 2007.

[9] J. H. Heo and M. C. Lee, "Building recognition using image segmentation and color features," *J. Korea Robotics Soc.*, vol. 8, no. 2, pp. 82-91, Jun. 2013.

[10] W. Zahng and J. Kosecka, "Localization based on building recognition," *IEEE Computer Soc. Conf.*, Jun. 2005.

[11] V. Vapnik, *The nature of statistical learning theory*, Springer, 1995.

[12] H. Trinh, D. N. Kim, and K. H. Jo, "Facet-based multiple building analysis for robot intelligence," *Mathematics and Computation*, vol. 205, no. 2, pp. 537-549, Nov. 2008.

[13] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM,* vol. 24, no. 6, pp. 381-395, Jun. 1981.

[14] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, "Speeded-up robust feature," *Computer Vision and Image Understanding*, vol. 10, no. 3, pp. 346-359, Jun. 2008.

[15] S. M. Smith and J. M. Brady, "Susan – a new approach to low level image processing," *Int. J. Computer Vision*, vol. 23, no. 1, pp. 45-78, May 1997.

[16] E. Rosten and T. Drummond, "Machine learning for high-speed corner detection," *Eur. Conf. Computer Vision*, pp. 430-443,

Graz, Austria, May 2006.
[17] L. M. J. Florack, B. M. T. H. Romeny, J. J. Koenderink, and M. A. Viergever, “General intensity transformations and differential invariants,” *J. Mathematical Imaging and Vision*, vol. 4, no. 2, pp. 171-187, May 1994.
[18] E. Dubrofsky, *Homography estimation*, Univ. of British COLUMBIA, Mar. 2009.
[19] M. M. Hossain, H. J. Lee, and J. S. Lee, “Fast image stitching for video stabilization using sift feature points,” *J. KICS*, vol. 39, no. 10, pp. 957-966, Oct. 2014.
[20] B. W. Chung, K. Y. Park, and S. Y. Hwang, “A fast and efficient haar-like feature selection algorithm for object detection,” *J. KICS*, vol. 38, no. 6, pp. 486-497, Jun. 2013.
[21] J. H. Hong, B. C. Ko, and J. Y. Nam, “Human action recognition in still image using weighted bag-of-features and ensemble decision trees,” *J. KICS*, vol. 38, no. 1, pp. 1-9, Jan. 2013.

박 시 영 (Si-young Park)

2015년 2월 : 광운대학교 전자공학과 학사
2015년 3월~현재 : 광운대학교 전자공학과 석사
<관심분야> 컴퓨터 비전, 영상 인식, 영상 신호 처리

안 하 은 (Ha-eun An)

2014년 2월 : 광운대학교 전자공학과 학사
2016년 2월 : 광운대학교 전자공학과 석사
2016년 3월~현재 : 광운대학교 전자공학과 박사
<관심분야> 영상통신, 영상인식, 디지털신호처리

이 규 철 (Gyu-cheol Lee)

2013년 2월 : 광운대학교 전자공학과 학사
2015년 2월 : 광운대학교 전자공학과 석사
2015년 3월~현재 : 광운대학교 전자공학과 박사
<관심분야> 3D 입체영상처리/압축, 스테레오 매칭, 영상 신호 처리

유 지 상 (Ji-sang Yoo)

1983년 2월 : 서울대학교 전자공학과 학사
1987년 2월 : 서울대학교 전자공학과 석사
1993년 5월 : Purdue Univ. EE, ph. D
1997년 9월~현재 : 광운대학교 전자공학과 교수
<관심분야> 웨이블릿 기반 영상처리, 영상압축, 영상인식, 비선형 신호처리