

# Multi Sparse Windows 기반의 TAD-Adaptive Census Transform 을 이용한 스테레오 정합 알고리즘

이인규, 문병인\*  
경북대학교 전자공학부  
e-mail : bihmoon@knu.ac.kr

## Stereo Matching Algorithm Using TAD-Adaptive Census Transform Based on Multi Sparse Windows

Ingyu Lee, Byungin Moon\*  
School of Electronics Engineering, Kyungpook National University

### 요 약

최근 3 차원 깊이 정보를 활용하는 분야가 많아짐에 따라, 정확한 깊이 정보를 추출하기 위한 연구가 계속 진행되고 있다. 특히 ASW(Adaptive Support Weight)는 기존의 영역 기반 알고리즘의 정확도를 향상시키기 위한 방법으로 많이 이용되고 있다. 그 중에서 ACT(Adaptive Census Transform)는 폐색 영역이나 경계 영역에서 정확도가 낮다는 단점이 있었다. 본 논문에서는 정확한 깊이 맵(depth map)을 추출하기 위해, 기존의 ACT 를 개선한 스테레오 정합 알고리즘을 제안한다. 이는 잡음에 강하고 재사용성이 높은 MSW(Multiple Sparse Windows)를 기반으로, TAD(Truncated Absolute Difference)와 ACT 두 개의 정합 알고리즘을 동시에 사용하여 폐색 영역과 물체의 경계 영역에서 정확도가 낮은 기존의 방법을 개선한다. Middlebury 에서 제공하는 영상을 사용한 시뮬레이션 결과는 제안한 방법이 기존의 방법보다 평균적으로 약 1.9% 낮은 에러율(error rate)을 가짐을 보여준다.

### 1. 서론

스테레오 정합이란 두 대 이상의 카메라를 사용하여, 동일한 시간에 서로 다른 위치에서 촬영된 2 차원 영상들로부터 3 차원 정보를 획득하는 기술이다. 두 대의 카메라로 촬영한 좌/우 영상에서 동일한 지점을 찾아, 동일한 지점 간의 시차(disparity)를 획득함으로써, 3 차원 거리 정보를 가지는 깊이 맵을 획득할 수 있다. 현재 스테레오 비전 시스템은 무인자동차, 로봇 비전, 감시시스템, 3 차원 재구성 등 다양한 응용분야에서 사용되고 있다.

3 차원 정보를 가지는 깊이 맵을 획득하기 위한 스테레오 정합 방법은 전역 정합 법(global matching), 지역 정합 법(local matching)으로 나눌 수 있다[1]. 전역 정합 법은 영상 전체의 정보를 이용하기 때문에 보다 정확한 깊이 맵을 획득할 수 있지만, 알고리즘이 복잡하고 연산량이 많다는 단점을 가지고 있다. 지역 기반 정합은 몇 개의 특징점을 기준으로 정합을 하는 특징 기반 정합과 윈도우 기반의 영역 기반 정합으로 다시 분류할 수 있다. 영역 기반 정합은 연산 과정이 간단하고 연산량이 적은 장점 때문에 효율적으로 깊이 정보를 얻을 수 있다는 장점을 가지고 있다. 일반적으로 영역 기반 정합에서는 조명 변화에 강건한 특성을 지니는 CT(Census Transform)이 주로 사용된다[2].

하지만 지역 정합 법은 전역 정합 법에 비해 정확한 깊이 맵을 획득하지 못한다는 한계를 지니고 있다. 이러한 한계를 극복하기 위해서 최근에 ASW 를 이용하는 연구가 진행되었다[3-5]. ASW 는 영역 가중치 윈도우(support-weight window) 내 중심 픽셀과의 컬러 유사성과 근접성 등을 이용한 영역 가중치(support weight)를 줌으로써 왜곡과 복잡한 물체로부터 강한 특성을 얻는 방법이다. 그 중에서도 ACT 는 영역 가중치 윈도우 내 픽셀들의 원 매칭 코스트(raw matching cost)를 구하기 위해 픽셀 마다 CT 윈도우를 설정하여 가중치 센서스 벡터(weighted census vector)를 생성한다[5]. 그렇기 때문에 기존의 ASW 보다 정확한 깊이 맵을 획득할 수 있다. 하지만 폐색 영역이나 물체의 경계에서는 정확도가 낮아 물체의 경계가 뚜렷하지 않다는 단점이 있다. 또한 잡음 형태의 오매칭이 되는 부분이 산발적으로 나타나는 단점을 가지고 있다.

본 논문에서는 기존의 ACT 보다 폐색 영역과 물체의 경계에서 정확한 시차 값을 획득할 수 있는 MSW 기반의 TAD-ACT 를 제안한다.

### 2. Adaptive Census Transform

ACT 는 영역 가중치 윈도우 내 픽셀의 원 매칭 코

\* 교신저자

스트를 구하기 위해 필요한 가중치  $W(p,q)$ 는 수식 (1) 처럼 CT 의 중심 픽셀과의 컬러 유사성과 근접성을 사용하여 구해진다[5].  $\Delta c$ 와  $\Delta g$ 는 중심 픽셀과의 색상 차이와 유클리드 거리(euclidian distance)의 값을 의미한다.  $\gamma_c$ 는 실험적으로 측정된 값이며,  $\gamma_p$ 는 영역 가중치 윈도우의 크기에 비례한다[3].

$$W(p, q) = e^{-\left(\frac{\Delta c + \Delta g}{\gamma_p}\right)} \quad (1)$$

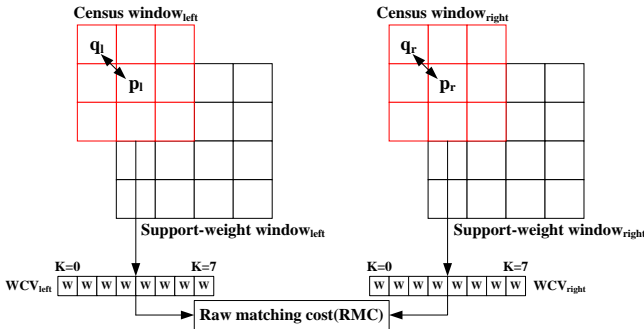
$(2r+1) \times (2r+1)$  크기의 CT 윈도우를 사용하는 ACT 의 원 매칭 코스트를 구하는 수식은 아래와 같이 정의된다.

$$\xi(p, q) = \begin{cases} -W(p, q) & \text{if } q \leq p \\ W(p, q) & \text{if } q > p \end{cases} \quad (2)$$

$$WCV(x, y) = \otimes_{i=-r}^r \otimes_{j=-r}^r \xi(I(x, y), I(x + i, y + j)) \quad (3)$$

$$RMC((x, y)_l, (x - d, y)_r) = \sum_{k=0}^{(2r+1)^2 - 2} |WCV_{l,k}(x, y) - WCV_{r,k}(x - d, y)| \quad (4)$$

수식 (2)에서  $\xi(p,q)$ 는 CT 윈도우 내 중심 픽셀  $p$  와 주변 픽셀  $q$  사이의 관계함수이며,  $W(p,q)$ 는 CT 윈도우 내 중심 픽셀과 주변 픽셀을 이용하여 구해진 가중치를 의미한다. 수식 (3)에서  $\otimes$ 는 일련의 가중치 벡터를 의미하며,  $WCV(x,y)$ 는 중심 픽셀을 기준으로 일련의 가중치 센서스 벡터를 생성하는 함수이다. 최종적으로 원 매칭 코스트는 수식 (4)처럼 좌/우 영상에서 설정된 CT 윈도우에 의해 생성된 가중치 센서스 벡터의 값을 각 인덱스  $k$  마다 AD(Absolute Difference)를 수행하여 나온 결과 값들을 모두 합하여 구해진다. 그림 1은 ACT의 원 매칭 코스트를 계산하는 과정에 대한 예시를 나타내고 있다.



(그림 1) ACT의 원 매칭 코스트 계산 과정

ACT의 최종 매칭 코스트는 수식 (5)처럼 좌/우 영상에서 설정된 영역 가중치 윈도우 내의 각 픽셀에 대한 영역 가중치와 원 매칭 코스트를 이용하여 구한 값들을 모두 합산하여 구한다.

$$MC_{final}(p, p_d) = \frac{\sum_{q \in N_p, q_d \in N_{p_d}} W_{support}(p, q) W_{support}(p_d, q_d) RMC(q, q_d)}{\sum_{q \in N_p, q_d \in N_{p_d}} W_{support}(p, q) W_{support}(p_d, q_d)} \quad (5)$$

수식 (5)에서  $p$  와  $p_d$ 는 각각 좌/우 영상에서의 영역 가중치 윈도우의 중심 픽셀을 의미하며,  $q$  와  $q_d$ 는 각각 좌/우 영역 가중치 윈도우에서 주변 픽셀을 의미한다.  $N_p$ 와  $N_{p_d}$ 는 영역 가중치 윈도우를 의미한다. 영역 가중치 윈도우의 중심 픽셀과 주변 픽셀의 컬러 유사성과 근접성을 이용한 가중치  $W_{support}$ 를 구하는 방법은 수식 (1)과 동일한 방법으로 구해진다.  $RMC$ 는 앞서서 구한 원 매칭 코스트를 의미한다.

최종적인 시차는 수식 (6)처럼 시차 추정 범위(disparity search range)만큼 계산된 최종적인 매칭 코스트들 중 가장 작은 매칭 코스트 값을 선택하는 방법인 WTA(Winner-Takes-All)를 이용하여 구해진다.

$$d = \operatorname{argmin}_{d \in \{d_{min}, \dots, d_{max}\}} MC_{final}(p, p_d) \quad (6)$$

그러나 기존의 ACT는 폐색 영역이나 경계 영역에서 매칭이 정확히 되지 않아, 경계가 뚜렷하지 않은 문제점을 가지고 있다. 또한 잡음 형태의 오매칭되는 부분이 산발적으로 나타나서, 전체적인 매칭률이 감소하게 된다. 이러한 문제를 해결하기 위해 본 논문에서는 MSW 기반의 TAD-ACT를 이용한 스테레오 정합 알고리즘을 제안한다.

### 3. 제안하는 MSW 기반의 TAD-ACT을 이용한 스테레오 정합 알고리즘

MSW 방법은 인접하는 SW(Sparse Window)를 두 개 이상을 종합하여 매칭 코스트를 추출하는 방법이다[6]. MSW 방법을 ACT에 적용한다면, WCV를 생성하기 위해 두 개 이상의 중심 픽셀을 사용하기 때문에 매칭 성능이 기존보다 향상될 수 있다[6].

본 논문에서 제안하는 MSW 기반의 TAD-ACT는 각 SW마다 중심 픽셀을 이용한 컬러 기반의 TAD와 ACT를 동시에 이용하여, 각 SW의 매칭 코스트를 구한다. 그리고 각 SW의 매칭 코스트를 모두 종합하여 최종적인 원 매칭 코스트를 추출한다.

먼저 TAD는 픽셀간의 색상 차이를 이용하는 것이며, CIELab 컬러 형식의 TAD의 수식은 (6)처럼 표현된다.  $T$ 는 폐색 영역으로 인해 매칭 코스트의 값이 커지는 것을 방지하기 위한 상수이다.

$$MC_{TAD,d}(I_l(x, y), I_r(x - d, y)) = \min(\sum_{c \in \{L, a, b\}} |I_{l,c}(x, y) - I_{r,c}(x - d, y)|, T) \quad (6)$$

ACT는 기존과 동일하게 영역 가중치 윈도우 내 픽셀마다 CT 윈도우를 설정하여, CT 윈도우 내 중심 픽셀과 주변 픽셀간의 상관 관계를 이용한다. 하지만 MSW를 기반으로 하기 때문에, SW들의 중심 픽셀로 이루어진 CW(Center Window)의 크기가  $a \times b$ 일 때, 각 SW의 WCV 수식은 아래와 같이 다시 표현된다.

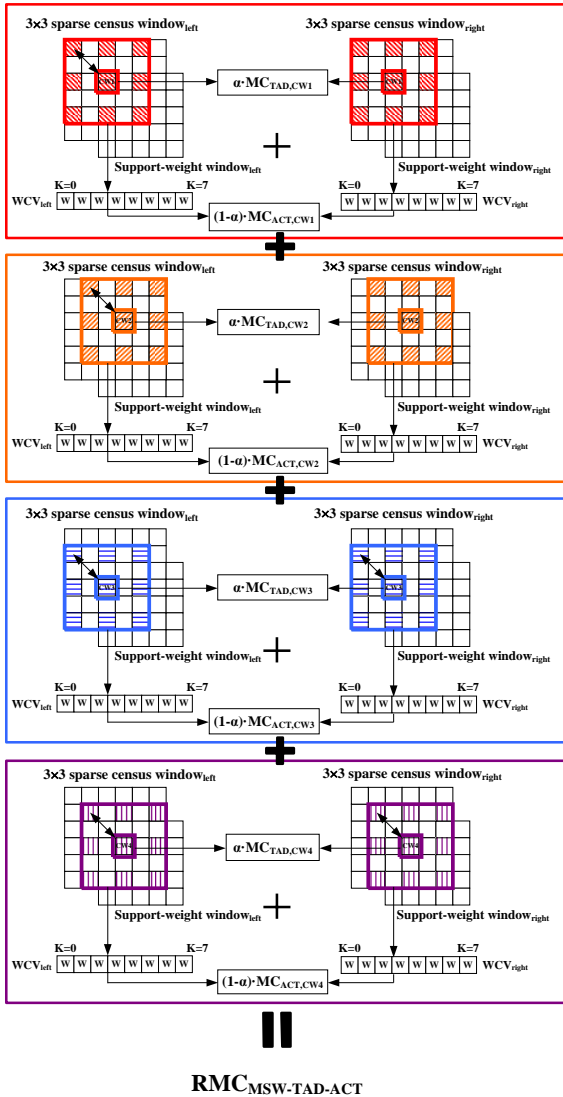
$$WCV_{MSW-ACT}(x, y) = \otimes_{i=-r}^r \otimes_{j=-r}^r \xi(I(x, y), I(x + i * a, y + j * b)) \quad (7)$$

제안하는 알고리즘은 앞서 설명한 두 개의 알고리즘을 각 SW 마다 동시에 사용하여 매칭 코스트를 추출한다. 그리고 각 SW 마다 구해진 매칭 코스트를 모두 합산하여 영역 가중치 윈도우 내 픽셀의 원 매칭 코스트를 구한다. 수식은 아래와 같이 표현된다.

$$RMC_{MSW-ACT-TAD}((x, y)_l, (x - d, y)_r) = \sum_u \sum_v \left( (1 - \alpha) \cdot MC_{ACT}(I_l(x + u, y + u), I_r(x - d + u, y + v)) + \alpha \cdot MC_{TAD}(I_l(x + u, y + v), I_r(x - d + u, y + v)) \right) \quad (8)$$

where  $u = \begin{cases} [-2/a + 1, 2/a] & \text{if } a \text{ is even} \\ [-(a-1)/2, (a-1)/2] & \text{if } a \text{ is odd} \end{cases}$   
 and  $v = \begin{cases} [-2/b + 1, 2/b] & \text{if } b \text{ is even} \\ [-(b-1)/2, (b-1)/2] & \text{if } b \text{ is odd} \end{cases}$

a 와 b 는 CW 의 x 축, y 축 크기를 나타낸다. 그림 2 는 2x2CW 를 사용할 때, MSW 기반의 TAD-ACT 예시를 나타낸다. 4 개의 3x3SW 각각의 매칭 코스트를 모두 종합하여 최종적인 원 매칭 코스트를 추출한다.



(그림 2) MSW 기반의 TAD-ACT 스테레오 정합 알고리즘 예시

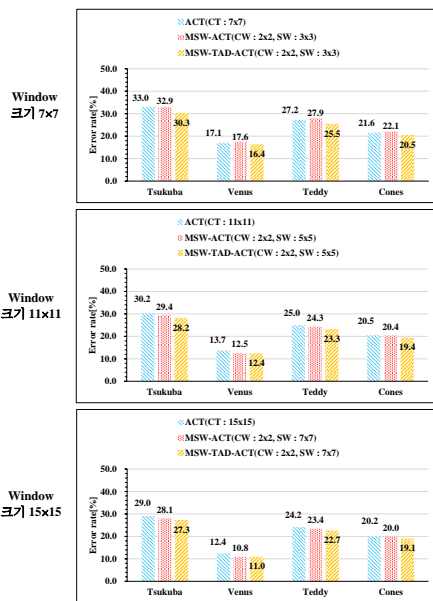
최종 매칭 코스트는 수식 (5)처럼 이전과 동일한 방법으로, 좌/우 영상에 설정된 영역 가중치 윈도우 내 각 픽셀의 영역 가중치와 원 매칭 코스트를 곱한 값을 모두 합산하여 구해진다.

기존의 ACT 는 폐색 영역이나 경계 영역처럼 값이 급변하는 영역에서 잘못된 매칭으로 인해 경계가 무너지는 단점이 있었다. 하지만 본 논문에서는 하나의 중심 픽셀을 사용하는 것이 아니라 여러 중심 픽셀을 사용하여 보다 정확성을 높이는 MSW 방식과 폐색 영역과 물체의 경계 영역에서 간단하면서도 좋은 매칭 성능을 보이는 TAD 를 같이 사용함으로써, 폐색 영역과 물체의 경계 영역에서 신뢰성 높은 시차를 추출한다. 또한 약간의 블러링을 유도하여 잡음 형태의 오매칭 되는 부분을 보정한다.

#### 4. 실험환경 및 결과

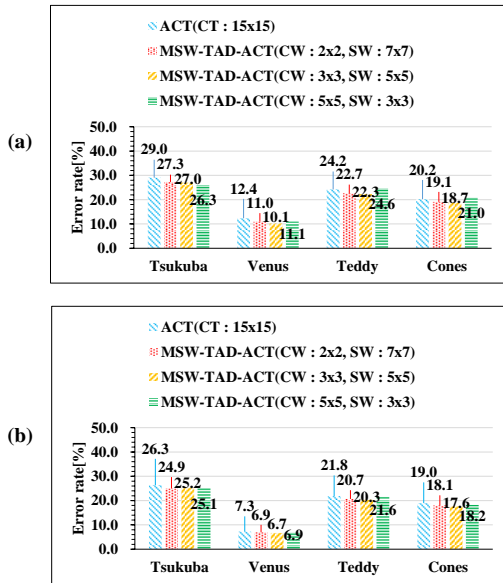
본 논문에서 제안하는 방법의 성능을 비교 평가하기 위하여 제안하는 방법과 기존의 방법들을 모두 OpenCV 라이브러리와 C/C++을 이용하여 모델링 하였다. 실험에는 middlebury 에서 제공하는 영상을 사용하였다[7]. 파라미터  $\gamma_c$  와  $T$  는 16, 40 으로 설정하였으며,  $\alpha$  는 실험적으로 구한 값인 0.1 로 설정하여 실험을 수행하였다. 성능 비교를 위해 middlebury 에서 제공하는 ground truth 영상과 추출한 깊이 맵 영상을 비교하여 전체적인 에러율을 측정하였다.

먼저 영역 가중치 윈도우 크기가 5x5 일 때, 기존의 ACT, MSW 기반의 ACT 및 MSW 기반의 TAD-ACT 의 에러율을 측정하였다. 그 결과 그림 3 처럼 기존의 ACT 를 그대로 사용하는 것보다, MSW 방법을 사용하는 것이 대체적으로 에러율을 감소시켜 준다는 것을 확인 하였다. 또한 제안한 방법이 기존의 방법 보다 평균적으로 에러율이 tsukuba 는 2.1%, venus 는 0.9%, teddy 는 1.6%, cones 은 1.1% 감소하였다.

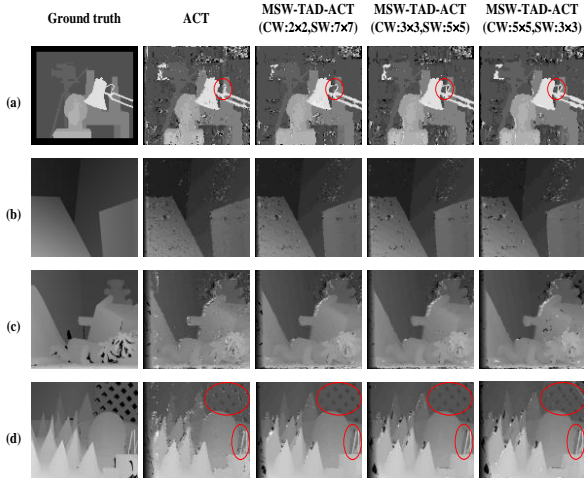


(그림 3) 윈도우 크기 별, ACT, MSW 기반의 ACT 및 MSW 기반의 TAD-ACT 의 성능 비교

또한 MSW 기반의 TAD-ACT 의 CW 변화에 따른 성능 분석을 위해 CW 의 크기를 다르게 하여 에러율을 측정하였다. 그 결과 그림 4 처럼 CW 의 크기는 3x3 일 때가 가장 에러율이 낮다는 것을 확인하였다. CW 의 크기가 3x3 과 5x5 를 비교 하였을 때, tsukuba 의 경우에는 에러율이 0.7% 감소하였지만, venus, teddy, cones 은 에러율이 각각 1%, 2.3%, 2.3% 증가하였다. 또한 그림 5 처럼 깊이 맵을 비교한 결과, 기존의 방법 보다 잡음 형태의 오차가 감소한 것을 확인할 수 있으며, 경계 영역에서 좀 더 정확한 시차 값을 출력하는 것을 확인 할 수 있었다. 그리고 적당한 크기의 CW 를 사용하게 되면 약간의 블러링 효과를 통해 매칭률이 향상되지만, CW 의 크기가 너무 커지게 되면 에러율이 높아지는 것을 확인하였다.



(그림 4) MSW 기반의 TAD-ACT 의 CW 크기 별 성능 비교: (a) 영역 가중치 윈도우 크기 5x5, (b) 영역 가중치 윈도우 크기 9x9



(그림 5) 영역 가중치 윈도우 크기가 5x5 일 때, 깊이 맵 결과: (a) Tsukuba, (b) Venus, (c) Teddy, (d) Cones

### 5. 결론

기존의 ACT 는 폐색 영역과 경계 영역에서 정확도가 낮은 동시에 점 형태의 오매칭이 발생하는 단점이 있었다. 이러한 단점을 개선하기 위해서 본 논문에서는 잡음에 강하고 재사용성이 높은 MSW 의 방식과 간단한 연산으로도 폐색 영역과 경계 영역에서 좋은 성능을 나타내는 TAD 를 융합함으로써, 기존보다 잡음에 강인하면서도 폐색 영역과 경계 영역에서 높은 정확성을 나타낼 수 있는 스테레오 정합 방법을 제안하였다. 실험 결과 기존의 방법에서 산발적으로 나타나는 점 형태의 오매칭 되는 부분을 제거하였으며, 그림 4 의 (a)에서 기존의 방법과 제안한 방법의 CW 크기가 3x3일 때, tsukuba, venus, teddy, cones 영상 각각의 에러율이 2.7%, 1.3%, 1.9%, 1.5% 감소하였다. 또한 폐색 영역과 물체의 경계 영역에서 정확히 매칭되지 않았지만, 제안한 방법은 폐색 영역과 경계 영역 모두에서 보다 정확히 매칭되는 것을 확인하였다.

### < Acknowledgement >

본 논문은 한국전자통신연구원(ETRI) 정부출연급 주요사업 ‘상황인지 스마트카를 위한 다중 센서 플랫폼 기술 개발’의 지원을 받아 수행되었음.

### 참고문헌

- [1] D. Scharstein and R. Szeliski, “A taxonomy and evaluation of dense two-frame stereo correspondence algorithms”, *International Journal of Computer Vision*, vol. 47, pp. 7-42, 2002.
- [2] R. Zabih and J. Woodfill, “Non-parametric local transforms for computing visual correspondence”, in *Proc. Third European Conference on Computer Vision (ECCV)*, Stockholm, Sweden, May 1994, pp. 151-158.
- [3] K.-J. Yoon and I. S. Kweon, “Adaptive support-weight approach for correspondence search”, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 4, pp. 650-656, Apr. 2006.
- [4] A. Hosni, M. Bleyer and M. Gelautz, “Secrets of adaptive support weight techniques for local stereo matching”, *Computer Vision and Image Understanding (CVIU)*, vol. 117, no. 6, pp. 620-632, Jun. 2013.
- [5] S. Perri, P. Corsonello and G. Cocorullo, “Adaptive Census Transform: A novel hardware-oriented stereovision algorithm”, *Computer Vision and Image Understanding (CVIU)*, vol. 117, no. 1, pp. 29-41, Jan. 2013.
- [6] K.-R. Bae, H.-S. Son, J. Hyun and B. Moon, “A census-based stereo matching algorithm with multiple sparse windows”, in *Proc. the International Conference on Ubiquitous and Future Networks (ICUFN)*, Sapporo, Japan, Jul. 2015, pp. 240-245.
- [7] *Middlebury Stereo Vision*. Available: <http://vision.middlebury.edu/stereo/>