

타원체 MODEL을 사용한 얼굴 영상의 시점합성에 관한 연구

학생회원 윤 나 리*, 종신회원 이 병 옥**

Viewpoint interpolation of face images using an ellipsoid model

Naree Yoon* *Student Member*, Byung-Uk Lee** *Lifelong Member*

요 약

영상 회의 시스템에서는 현실감을 높이기 위하여 대화 상대방과 시선을 맞추는 eye contact을 만들어 주는 것이 중요하다. 이를 위하여 영상 기반 방법과 모델 기반 방법 등의 시점합성기법이 연구되고 있다. 전자는 계산이 간단하나 부정확하고, 후자는 잡음에 강하고 정확한 영상을 만들어 내지만 시간이 많이 걸리는 단점이 있다. 본 논문에서는 얼굴 영상을 타원으로 근사화시키고 이를 3차원 공간에서 타원체로 모델링하여 임의의 가상 카메라의 위치에서 영상을 합성하는 방법을 제안하였다. 여러 가지 실험을 통하여 제안한 방법이 계산량이 적으면서도 성능이 우수함을 입증하였다.

Key Words : Viewpoint interpolation, Eye contact, Ellipsoid head model

ABSTRACT

To establish eye contact in video teleconferencing, it is necessary to synthesize a front view image by viewpoint interpolation. We can find the viewing direction of a user, and interpolate an image seen from that viewpoint, which will result in a face image observed from the front. There are two categories of previous research: image based method and model based method. The former is simple to calculate, however, it shows limited performance for complex objects. And the latter is robust to noise while it is computationally expensive. We propose to approximate face images as ellipses and match them to build an ellipsoid and then synthesize a new image from a given virtual camera position. We show that it is simple and robust from various experiments.

I. 서론

최근 인터넷의 발달과 고속 통신망의 보급으로 인하여 업무 및 개인용으로 영상 회의 혹은 영상 전화를 사용하는 일이 보편화되고 있다. 그러나 영상회의시 사용자가 화면상의 다른 사용자를 바라보아도 카메라는 화면 밖에 위치하기 때문에, 카메라에 촬영된 영상은 다른 곳을 바라보고 있는 느낌을

준다. 정면을 바라보는 영상은 사용자가 카메라를 바라볼 때만 얻을 수 있다. 그러나 물리적으로 카메라를 화면의 다른 사용자의 얼굴 영상 뒤에 두는 것은 불가능하다. 영상회의에서 상대방과 시선을 맞추는 eye contact은 현실감을 충족시키기 위한 매우 중요한 요소 중의 하나이다. Eye contact을 만족시키기 위해 왼쪽과 오른쪽 혹은 위와 아래에서 얻은 영상을 합성하여 정면을 바라보는 영상을 만들어

※ 본 연구는 정보통신부 및 정보통신연구진흥원의 대학 IT연구센터 지원사업의 연구결과로 수행되었음.

* 이화여자대학교 정보통신학과 영상처리연구실, ** 이화여자대학교 정보통신학과 교수

논문번호 : KICS2005-05-198, 접수일자 : 2005년 5월 16일, 최종논문접수일자 : 2007년 5월 9일

낼 수 있다¹¹. 이와 같은 용도에서는 실시간으로 영상을 합성해야 하므로 그 알고리즘은 간단하면서도 오차에 강해야 한다.

기존의 연구 분야는 크게 두 가지로 나눌 수 있다. 영상 기반 방법^{[2][3][8][9]}과 모델 기반 방법^{[10][12]}이다. 영상 기반 방법은 여러 시점에서 취득한 영상을 이용하여 2D 영상 공간에서 다른 시점의 영상을 합성해낸다. 그러므로 상대적으로 간단하고 빠르게 실시간이 요구되는 응용에 적합하다. 반면, 모델기반 방법은 취득된 영상들로부터 3D 모델을 만들어 낸 후 카메라 모델을 적용하여 2D 영상을 만든다. 이 방법은 현실감 있는 영상을 만들어 내는 3D 모델을 만들어 내는 데에 시간이 많이 소요되는 단점이 있다. 이 외에도 거리 영상으로부터 다시점 영상을 합성하는 방법도 있다¹³.

본 논문에서는 얼굴을 타원체로 간단히 모델링하여 계산량이 적으면서도 오차에 강인한 방법을 제안한다. 먼저 영상의 얼굴을 타원으로 가정하고 두 개의 영상의 타원으로부터 3 차원 타원체를 만들어 낸다. 그리고 만들어진 타원체를 촬영한 영상과 가상카메라에 투사시켜서 대응점을 찾는다. 촬영된 영상의 대응점들의 밝기를 평균하여 가상 카메라 영상의 밝기를 얻을 수 있다. 이 방법을 사용하여 간단하면서도 만족할 만한 합성 영상을 얻을 수 있다.

II. 기존의 시점 합성 연구

스테레오 영상의 기법을 사용하여 영상 공간에서 시점 합성을 하는 방법이 연구되어왔다. 두 개의 촬영된 영상에서 대응점을 찾아 disparity를 계산하고, 이를 사용하여 새로운 가상 영상 위치에서의 disparity를 계산하여 영상을 합성할 수 있다. 이 경우 가상 카메라의 초점의 위치는 두 개의 카메라의 초점을 연결하는 직선상에 있어야만 하는 제약이 있다. 스테레오 영상에서 disparity를 찾는 데에는 여러 방법이 연구되었다^{[2][3][8]}. Projective grid method는 가상의 카메라의 영상 한 점과 초점을 잇는 직선상에서 최적의 위치를 찾는다^{[11][12]}. 이 직선상의 점을 각각 좌, 우 영상에 투사하여 두 영상의 밝기가 가장 비슷한 위치가 실제 물체의 위치로 가정한다. 이 방법은 조명조건이나 영상잡음, 영상의 텍스처 부족, 카메라의 불일치 등으로 인하여 잘못된 위치를 찾을 가능성이 높고, 이는 심각한 오류를 유발할 수 있다.

얼굴 모델링 방법은 스테레오 영상기법^{[2][14]}을 적용한 방법과 personalized face model^[15]을 사용한 방법이 있다. Personalized face model을 이용한 3D 스테레오 머리 추적기는 두 개의 시점 사이의 초기 대응점을 찾는 데 이용된다. 더 자세한 대응점이 template matching과 feature matching으로 추가된다. 결국, 모든 대응점 정보는 시점합성을 위해 합쳐지고 view morphing방법을 이용하여 시점합성을 수행한다. 이 방법은 합성된 시점의 정확도와 강인함을 크게 향상시킨다. 그러나 이 방법은 계산이 매우 복잡하여 실시간 응용에는 적합하지 않다^[10].

III. 얼굴 모델을 적용한 제안된 방법

영상의 밝기만을 사용하여 대응점을 찾는 방법은 국부적인 정보만을 사용하므로 잡음에 민감하다. 본 논문에서는 전역적인 정보를 사용하여 얼굴의 기하학적 모델을 생성하였다. 따라서 간단하고 정확하게 카메라 영상들을 정합하였다. 첫 단계로 평면 모델을 사용하였으나 얼굴 옆면에서의 오차가 심하여 세 개의 평면을 사용하였으나 만족할 만한 결과를 얻지 못하였다. 실험 결과는 다음 절에 보였다. 이를 개선하기 위하여 인간의 얼굴에 타원체 모델을 적용하여 간단하면서도 우수한 영상을 합성하였다. 먼저 영상으로부터 얼굴을 분할하고 그 윤곽선을 타원으로 모델링한다. 영상에서 얼굴을 분리하고 추적하는 알고리즘은 실시간으로도 구현가능하다^{[16][17]}. 그 후에 그림 1에 보인 바와 같이 두개의 영상 타원의 장축과 단축의 끝점을 정합하여 타원체를 만들어 낸다. 가상의 카메라는 3차원 공간상의 어디에나 위치시킬 수 있고, 그 자리에서 촬영한 영상을 합성해 낼 수 있다. 먼저 가상의 카메라의 초점으로부터 가상의 영상 평면 위의 화소를 통과하는 직선을 계산한다. 이 직선이 타원체 얼굴 모델과 만나면 그 공간상의 만나는 점을 촬영된 두 영상평면으로 투사하고 그 투사된 점의 밝기의 평균을 구한다. 그 후 이와 같은 과정을 합성할 영상의 각 화소에서 반복한다. 실제로 사람의 머리는 타원체보다 훨씬 복잡하나 간단한 머리 모델로 만족할 만한 얼굴 영상을 얻을 수 있다. 본 논문에서 제안한 방법은 2차원 영상공간에서의 대응점을 찾는 것이 아니고 3차원 대응관계를 이용하므로 잡음에 강인하다. 또한 3차원 얼굴모델로 타원체를 사용하여서 모델 생성이 간단하여 계산량이 적다.

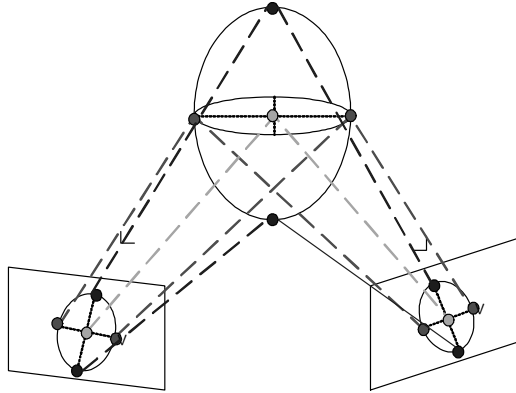


그림 1. 타원체 얼굴 모델링
두개의 얼굴 영상을 타원으로 모델링하고 이를 정합하여 3차원 타원체를 만든다.

IV. 실험 결과

본 절에서는 기존 방법인 projective grid space 방법의 결과를 보이고 얼굴 타원체 모델링의 전 단계인 평면 모델의 결과를 보인다. 실험영상은 두 개의 디지털 카메라로 동시에 촬영하였다. 그림 2 (a)와 (b)는 카메라로부터 촬영된 영상을 보여준다.

카메라는 calibration되었기 때문에, 카메라의 내부변수와 외부변수는 알고 있다고 가정한다. [18][19][20] 두 개의 영상은 카메라 특성의 불일치와 조명조건의 불일치로 인하여 밝기가 다를 수 있다. 이를 보정하기 위하여 얼굴부위의 평균과 표준편차를 이용

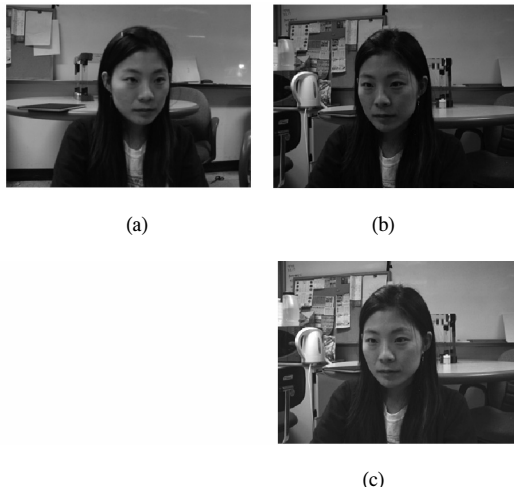


그림 2. (a) 왼쪽 영상 (b) 오른쪽 영상 (c) 밝기가 보정된 오른쪽 영상

하여 좌영상과 우영상의 밝기를 식 (1)을 사용하여 영상 평균과 표준편차를 맞추어 준다.

$$I_R'(u, v) = \frac{\sigma_L}{\sigma_R} (I_R(u, v) - \mu_R) + \mu_L, u, v \in I_R \quad (1)$$

위 식에서 μ_R 과 μ_L 는 좌영상과 우영상의 밝기 평균이며, σ_R 와 σ_L 는 좌영상과 우영상의 표준편차이다. 그림 2(c)는 그림 2(b)의 보정된 영상이다.

4.1. Projective grid space

그림 3은 projective grid space 방법의 결과를 보여준다. 그림 4에서 가상카메라의 초점과 합성영상의 한 화소를 연결하는 직선에 따른 대응점 밝기 차이를 보여준다. 여기에서 밝기가 동일한 점이 여러 곳이 발생한다. 따라서 이러한 방식으로 구한 밝기에 오류가 발생할 수 있다. 또한 대응점의 밝기는 영상의 잡음이나 카메라의 노출의 차이 등으로 인하여 오차가 발생할 수 있다. 그러므로 이 방법은 그림 3과 같이 오동작할 가능성이 높다.



그림 3. Projective grid space method의 결과 영상

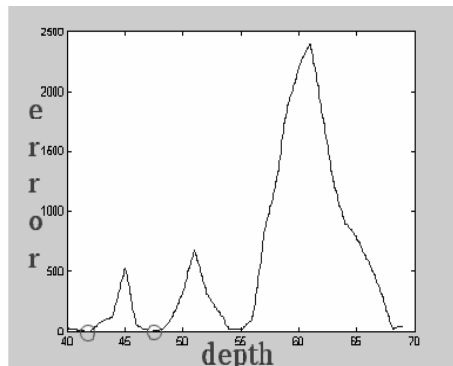


그림 4. Projective grid space에서 밝기가 동일한 지점들

4.2. 평면 모델링

4.2.1. 평면 1개를 이용한 경우

얼굴까지의 거리가 비슷하므로 얼굴을 평면으로 가정해 보았다. 얼굴의 평면 모델을 눈의 두 점과 코 끝점을 이용하여 정의한다. 얼굴 외의 부분은 위와 같이 깊이에 따른 밝기를 비교하여 interpolation 한다.

그림 5의 결과를 보면 실제로 얼굴의 많은 부분은 평면에 가까워서 영상이 정상적으로 보이지만 평면으로부터 많이 벗어나는 부분은 제대로 interpolation되지 않는다.

4.2.2. 3개의 plane 모델

얼굴을 한 개의 평면으로 모델링한 경우 얼굴 옆면에서 많은 오차가 발생하였다. 이를 줄이기 위하여 얼굴을 3개의 평면으로 모델링하여 실험하였다.

볼 부분의 평면식을 계산하는 것이 중요하나 여기서는 영상이 좌, 우로 많이 떨어져 있기 때문에 occlusion부분이 크므로 실제 왼쪽과 오른쪽 볼의 평면 식을 알아내기가 어렵다. 이와 같이 카메라가 수평방향으로 떨어져 있는 경우에는 두 영상간의 대응점을 찾기 어려우므로 카메라를 상하 방향으로 배치하면 얼굴의 대응점이 가려진 부분이 적어서 가운데 영상을 만들기가 용이하다.

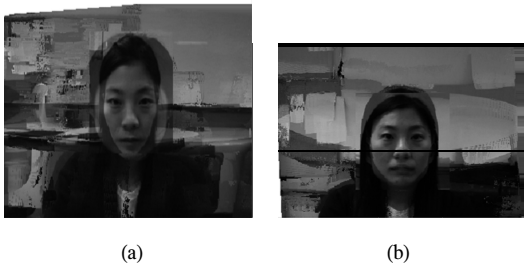


그림 5. 평면으로 복원한 영상 (a) 좌우영상을 이용한 합성 영상 (b) 상하 영상을 이용한 합성 영상

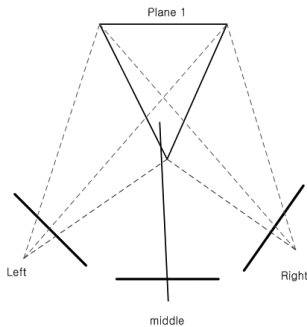


그림 6. 평면 세 개로 모델링하여 복원한 영상

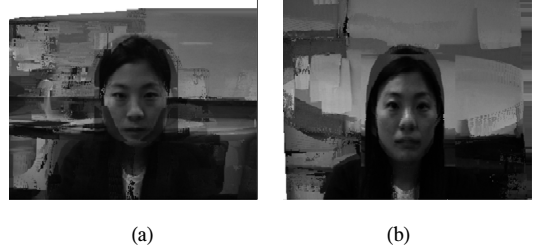


그림 7. 3개의 평면으로 복원한 영상 (a) 좌우 영상을 이용한 합성 영상 (b) 상하 영상을 이용한 합성 영상

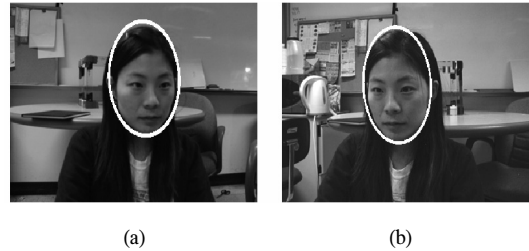


그림 8. 얼굴의 타원체 모델. (a) 왼쪽 영상 (b) 오른쪽 영상

4.3. 얼굴의 타원체 모델링

먼저 왼쪽과 오른쪽 영상으로부터 얼굴과 머리를 분할한다. 그 다음 전 처리 과정으로 얼굴 부분의 평균 밝기와 표준편차가 같아지도록 밝기보정을 한다. 그 후에 그림 8과 같이 얼굴의 윤곽선을 이용하여 최소자승오차 (MMSE: minimum mean square error) 타원을 계산한다. 그 후 두 영상에서 타원의 장축과 단축의 끝점을 정합시켜 타원체를 만들어 낸다. 카메라 calibration과정을 통하여 카메라의 내부변수와 외부변수를 알고 있으므로 타원의 윤곽선 점들과 초점을 연결하는 두 직선의 최소자승오차 교점을 얻을 수 있다. 그 후 임의의 가상 카메라를 위치에서 영상을 합성한다. 영상회의에 나타난 대화 상대방의 눈에 가상카메라를 위치시킨다면 정면을 바라보는 영상을 얻을 수 있다. 사용자가 주시하는 방향을 검출하는 방법에 관한 연구도 최근에 발표된 바가 있다^{[11][21][22]}. 다양한 가상의 위치에서 합성 영상을 얻은 결과를 그림 9에 보였다. 왼쪽과 오른쪽 입력영상을 이용하여 다섯 개의 가상 카메라의 위치, 즉 가운데, 위, 아래, 왼쪽 근처, 오른쪽 근처에서 보여질 합성 영상을 보였다. 모든 합성된 영상은 자연스럽게 보이고 다른 위치에서 보는 것과 같은 느낌을 준다. 3차원 모델을 적용하였기 때문에, 원하는 어느 위치에서도 영상을 만들어 낼 수 있다. 그러므로 움직이는 위치에서 navigation하는

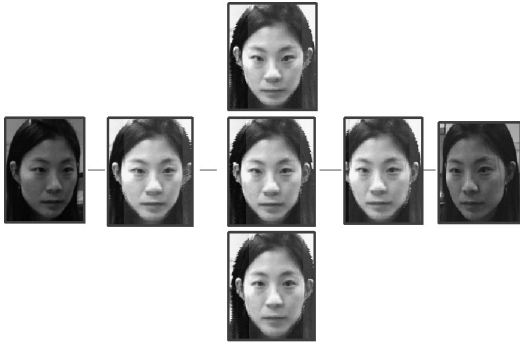
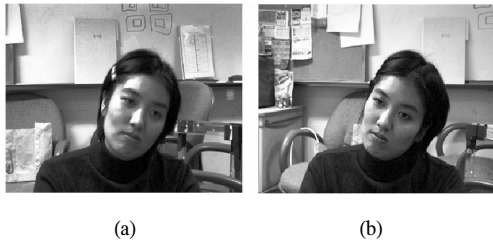


그림 9. 주어진 왼쪽, 오른쪽 영상을 이용해 만들어 낸 영상들



(a) (b)

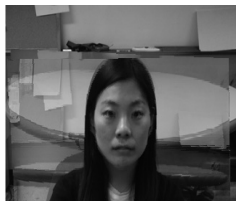


(c)

그림 10. 고개를 기울인 경우 (a) 왼쪽 영상(b) 오른쪽 영상 (c) 합성된 영상

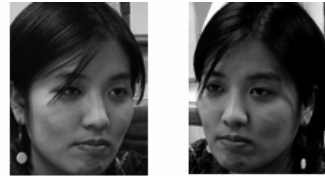


(a) (b)

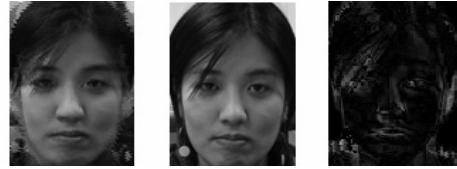


(c)

그림 11. (a) 위 영상 (b) 아래 영상 (c) 합성된 영상



(a) (b)



(c) (d) (e)

그림 12. (a) 왼쪽 영상 (b) 오른쪽 영상 (c) 타원체 얼굴 모델을 이용해 합성된 영상 (d) 정면에서 촬영한 영상 (e) Error image

영상도 만들어 낼 수 있다. 그림 10에서 보듯이 제안된 방법은 물체의 이동이나 머리의 이동 등에 강인하다.

카메라를 위, 아래로 위치시키는 경우에는 좌, 우로 위치시킨 경우에 비교해 보았을 때 영상간의 occlusion을 줄일 수 있다. 그림 11은 위, 아래로 카메라를 위치시킨 경우의 실험결과를 보여준다.

합성된 정면 영상과 정면에서 촬영한 영상간의 차이를 살펴보았다. 그림 12 (d)는 카메라로 촬영한 정면 실사영상이고, (c)는 좌영상 (a)와 우영상 (b)를 사용하여 (d)의 위치에서 합성된 영상이다. 합성된 영상은 정면에서 촬영한 실사 영상과 비슷하여 자연스럽게 보이는 것을 확인할 수 있었다. 실사 영상과 합성영상의 차이영상 (e)에서 윤곽선의 약간의 오차만 보인다.

V. 결론

본 논문에서는 두 개의 영상을 이용하여 임의의 가상의 위치에서 바라본 영상을 만들어 내어 영상 회의 동안 eye contact을 가능하게 하는 방법을 제안하였다. 촬영한 영상의 얼굴을 타원으로 모델링하고 두 장의 영상의 타원으로부터 3차원 공간에서 타원체로 머리를 모델링한다. 타원체를 가상 영상 평면으로 투사하여 영상 화소의 대응점을 찾는다. 이 방법은 기존의 영상기반 방법보다 잡음에 강인하다. 또한 3차원 모델이 타원체로 기존의 3차원 메쉬 모델에 비하여 훨씬 더 간단하므로 계산량이 획기적으로 줄어들어서 실시간 구현에 적합하다. 또

한 이 방법은 영상 기반 방법과 같이 촬영된 영상의 초점을 있는 직선에 한정되지 않기 때문에 촬영하는 가상 카메라위치 선정이 자유롭다.

본 논문의 방법은 얼굴 부분에만 제한되어 있고 배경 영상은 만들어 내지 않는다. 그러므로 자연스러워 보일 수 있는 배경영상을 빠르게 만들어 낼 수 있는 방법에 대해 고안할 필요가 있다. 또한 정확한 eye contact을 위하여 eye gaze algorithm^{[9][21][22]}을 적용하여 주시 방향을 구하여 가상카메라의 위치와 방향을 정할 필요가 있다.

참 고 문 헌

- [1] D. W. Hanson, "Eye typing off the shelf," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 2, pp. 159- 164, Washington D.C., U.S.A., Jul. 2004.
- [2] M. Z. Brown, D. Burschka, G. D .Hager, "Advanced in computational stereo," *IEEE transaction on pattern analysis and machine intelligence*, Vol. 25, No. 8, pp. 993-1008, Aug. 2003.
- [3] P. Aschwanden and W. Guggenbuhl, "Experimental Results from a comparative study on correlation-type registration algorithm," *Robust Computer Vision, Forstner and Ruviedel, Eds.*, pp. 268-289, Wickmann, 1993.
- [4] V. Venkateswar and R. Chellappa, "Hierarchical Stereo and Motion Correspondence Using Feature Groupings," *Int'l J. Computer Vision*, Vol. 15, pp. 245-269, 1995.
- [5] S. Birchfield and C. Tomasi, " Multiway cut for stereo and motion with slanted surfaces," *Proc. Int'l Conf. Computer Vision*, Vol. 1, pp. 489-495, 1999.
- [6] S. Birchfield and C. Tomasi, "Depth discontinuities by Pixel-to-Pixel Stereo," *Proceedings of the Sixth IEEE international Conference on Computer Vision, Mumbai, India*, pp. 1073-1080, Jan. 1998.
- [7] C. Tomasi and R. Manduchi, "Stereo without Search," *Technical Report STAN-CS-TR-95-1543*, Stanford Univ., 1995.
- [8] T. H. Cormen, C. E. Leiserson, and R. L. Rivest, *Introduction to Algorithms*. New York, McGraw-Hill, 1990.
- [9] 박창주, 고정환, 김은수, "스테레오 영상 보정 알고리즘에 기반한 새로운 중간시점 영상합성 기법," *한국통신학회논문지*, Vol. 29, No. 5C, pp. 632-641, May. 2004.
- [10] R. Yang, Z. Zhang, "Eye gaze correction with stereovision for video-teleconferencing," *IEEE transaction on Pattern Analysis and Machine Intelligence*, Vol. 26, No. 7, pp. 956-960, Jul. 2004.
- [11] H. Saito, M. Kimura, S. Yaguchi, N. Inamoto, "View interpolation of multiple cameras based on projective geometry," *IPSJ Transactions on Computer Vision and Image Media*, Vol.43, No. SIG 11, 2002.
- [12] J. Yin, and J. R. Cooperstock, "Improving depth maps by nonlinear diffusion," *12th International Conference on Computer Graphics, Visualization and Computer Vision*, Feb. 2004.
- [13] 장세훈, 한충신, 배진우, 유지상, "깊이 정보를 이용한 실시간 다시점 스테레오 영상 합성," *한국통신학회논문지*, Vol. 30, No. 4C, pp. 239-246, Apr. 2005.
- [14] A. M. K. Siu and R. W. H. Lau, "Image Registration for Image-Based Rendering," *IEEE Trans. on Image Processing*, Vol. 14, No. 2, pp. 241-252, Feb. 2005
- [15] Z. Liu, Z. Zhang, C. Jacobs, and M. Cohen, "Rapid modeling of animated faces from video," *Journal of Visualization and Computer Animation*, Vol. 12, No. 4, pp. 227-240, 2001.
- [16] M. Kass, A. Witkin, and D. Terzopoulos, "Snake : Active Countour models," *Int'l J. Computer Vision*, Vol. 1, No. 4, pp. 321-331, 1987.
- [17] M. H. Yang, D. Kriegman, N. Ahuja, "Detecting Faces in Images: A Survey," *IEEE Trans. on Pattern Analysis and Machine Vision*, Vol. 24, No. 1, pp. 34-58, Jan. 2004.
- [18] R. Tsai, "A versatile Camera Calibration

