

İşitsel ve Görsel Uzaktan Bulunma Sisteminin Tasarımı ve Uygulanması

Rabiye KARALI

İzmir Yüksek Teknoloji Enstitüsü
Elektrik-Elektronik Mühendisliği Bölümü
rabiyeKarali@std.iyte.edu.tr

Bekir Berker TÜRKER

İzmir Yüksek Teknoloji Enstitüsü
Elektrik-Elektronik Mühendisliği Bölümü
bekirturker@std.iyte.edu.tr

Şevket GÜMÜŞTEKİN

İzmir Yüksek Teknoloji Enstitüsü
Elektrik-Elektronik Mühendisliği Bölümü
sevkettumustekin@iyte.edu.tr

Özet

Görsel ve işitsel bir sanal gerçeklik sistemi geliştirilmesine dayanan bu çalışmada, sanal seyahat, uzaktan eğitim, uzaktan ameliyat, tele-konferans gibi birçok uygulamaya uyarlanabilecek şekilde kullanıcılarda uzakta bulunma algısı yaratılması hedeflenmiştir. Bu amaçla, ambisonik çevresel ses düzeni modeli ve tümyönlü video işleme teknikleri kullanılmıştır. Kullanıcının baş hareketleri baş takip sensörü sayesinde 3 serbestlik derecesiyle algılanarak gerçek zamanlı işlenen ses ve video, takılabilir video gözlüğü ve kulaklığa aktarılmaktadır. Uzak ortamda kaydedilen 4 kanallı ses ve tümyönlü video, kullanıcı tarafındaki sensörler aracılığıyla elde edilen yön bilgilerinin kullanımıyla işlenir. İşlenen ses akışı 2 kanallı kullanıcı kulaklığına, işlenen görüntü ise video gözlüğüne gönderilerek kullanıcının uzak ortamda bulunma hissini yaşaması sağlanır. Önceden kaydedilmiş ses ve video verisinin eşzamanlı olarak kullanıcı hareketleri ile uyumlu bir şekilde oluşturulması için geliştirilen bu uzaktan bulunma sistemi, kaydedilen ortamların yanı sıra gerçek zamanlı canlı aktarım uygulamalarında da kullanılabilir.

Anahtar kelimeler: uzaktan bulunma, yapay gerçek, ambisonik, tümyönlü

1. Giriş

Uzak bir ortamda bulunma hissini amaçlanan uygulamaya uygun güvenli bir ortamda bulunan bir kullanıcıya aktarılması, gittikçe gelişmekte olan sensör ve ekran teknolojilerinin kullanımı ile mümkün hale gelmiştir. Uzaktan bulunma teknolojisi, önceden kaydedilmiş ya da akış halindeki ses/video kullanılmasıyla kişinin sanal veya gerçek olan 3 boyutlu bir ortamda bulunma hissini yaşamasını mümkün kılmaktadır. Bu amaçla, çalışma gerçekleştirilirken ambisonik çevresel ses sistemi tekniği kullanılarak yönlü mikrofonlar ile ses kaydı alınmış olup tümyönlü yüksek çözünürlüklü (HD) kamera sayesinde video kaydı yapılmıştır. Kullanıcının önceden kaydedilmiş bu ortamda bulunma hissini yaşayabilmesi için ise gözlük şeklindeki görüntüleme cihazı ve kulaklık kullanılmıştır.

Literatürde, görüntüleme cihazının görsel açısını

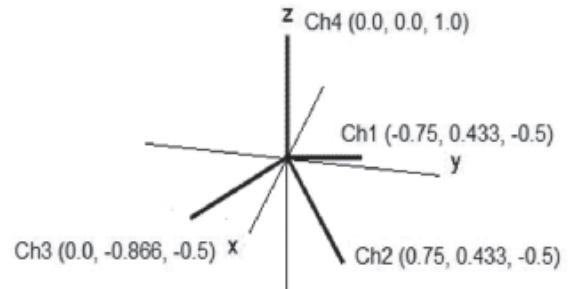
arttırmak [1] vb. çalışmaların yer alması görsel uzaktan bulunma sistemine katkı sağlamıştır. Internet aracılığıyla görsel uzaktan bulunma [2], tümyönlü HD kamera ile uzaktan bulunma sistemi [3], gerçek zamanlı görünüm bağımlı görüntü üretme [4] literatürde yer alan benzer çalışmalardır. Bu çalışmanın özgünlüğü, hem işitsel hem de görsel uzaktan bulunma sisteminin bir arada sunulmasına bağlıdır. Üstelik önceden kaydedilmiş ortam uygulamalarında sağlanan başarının canlı akış uygulamalarında da sağlanması beklenmektedir. İşitsel ve görsel olmak üzere iki ayrı kategoride çalışılıp sonradan birleştirilen uzaktan bulunma sisteminin detayları bir sonraki kısımda anlatılacaktır.

2. Yöntem

2.1. İşitsel Uzaktan Bulunma Sistemi

2.1.1. 4-Kanallı Ses Kaydı ve Vektör Modeli

İşitsel uzaktan bulunma sisteminde, ambisonik çevresel ses sistemi 3 boyutlu sesin algılanmasında önemli rol oynamaktadır. Birinci derece ambisonik çevresel ses sistem modeli, B-formatı [5][6], alınan 3 boyutlu sesin yönlü bilgisini 4 kanala kodlar. Bu model, 4 tane farklı yönlü işitsel mikrofonun dörtyüzlü (tetrahedron) yapının uç noktalarına yerleştirilmesini gerektirir. Her mikrofon bir kanalı ifade eder ve bulunduğu koordinat noktasına uygun birim vektör ile ifade edilir. Kanalların kartezyen koordinat sistemdeki yerleri (x,y,z) Şekil 1’de gösterilmiştir. Tetrahedron köşelerine yerleştirilen mikrofonlar yönlü mikrofon olduğundan bu yapıda kaydedilen ses, kaynak yönü bilgisini de saklamaktadır.



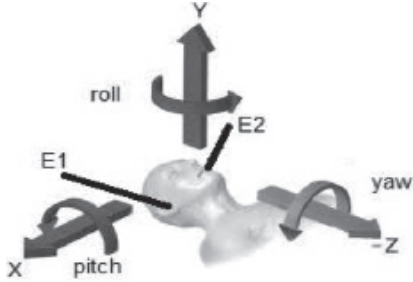
Şekil 1: 3-boyutlu uzayda kanal vektörleri.

2.1.2. Kulak Vektörlerinin Modellenmesi

Bu özelliği gerçekleştirmek üzere, üç boyutlu koordinatlarda insan kulağını başı merkez olacak şekilde modellemek gerekir. İki tane kulak vektörü, kanal vektörlerine uygun olan birim vektörlerle temsil edilebilir. Sağ ve sol kulak vektörlerinin kartezyen koordinatları sırasıyla şu şekildedir:

$$E1o: (0.866, 0, 5, 0) \quad (1)$$

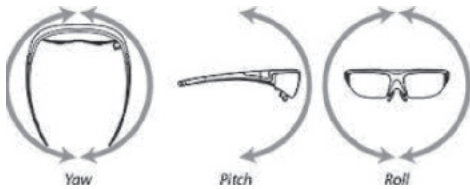
$$E2o: (-0.866, 0, 5, 0) \quad (2)$$



Şekil 2: 3-boyutlu uzayda: E1 sağ kulak vektörü ve E2 sol kulak vektörü.

2.1.3. Hareket Takibi ve Rotasyon Matrisi

Uzaktan bulunma hissini yaşayabilmek için kullanıcının kulak pozisyonlarının 3 serbestlik derecesindeki açı bilgileri (yaw (φ), pitch (Φ), ve roll (θ)) açıları sensör aracılığıyla alınmalıdır. Bu sensör, video gözlüğünde yer alır. Kulak vektörlerinin pozisyonları, sensörden yeni veri geldikçe yeniden hesaplanmalıdır. Bu hesaplama için, 3 boyutlu rotasyon matrisleri [7] Ryaw, Rpitch ve Rroll kullanılır. Üç boyutlu uzayda gerçekleşen hareketler ise rotasyon parametrelerinin kombinasyonu ile modellenir.



Şekil 3: Rotasyon parametreleri: yaw (sağa sola hareket), pitch (yukarı aşağı hareket), roll (omuzlara doğru hareket)

$$R_{yaw} = \begin{bmatrix} \cos\varphi & \sin\varphi & 0 \\ -\sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (3)$$

$$R_{pitch} = \begin{bmatrix} \cos\Phi & 0 & \sin\Phi \\ 0 & 1 & 0 \\ \sin\Phi & 0 & \cos\Phi \end{bmatrix} \quad (4)$$

$$R_{roll} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{bmatrix} \quad (5)$$

$$R_{total} = R_{yaw} R_{pitch} R_{roll} \quad (6)$$

$$E1 = R_{total} E1_0 \quad (7)$$

$$E2 = R_{total} E2_0 \quad (8)$$

E1o, E2o sağ ve sol kulak vektörlerinin ilk pozisyon değerleridir. E1, E2 ise yeniden hesaplanmış kulak vektörüdür.

Kanal karıştırma işlemi (mixing), her işitsel kanalın sağ ve sol kulak vektörleri üzerine ayrı ayrı iz düşürülmesi ile gerçekleştirilir. İç çarpım [9] hesaplamaları kullanılarak sesler ağırlıklandırılır ve her kulak üzerinde karıştırılır.

$$S1_{ağırlık}[n] = E1 \cdot S_{kanal}[n] \quad (9)$$

$$S2_{ağırlık}[n] = E2 \cdot S_{kanal}[n] \quad (10)$$

$$n = \{1, 2, 3, 4\}$$

$S_{kanal}[n]$, n'inci ses kanalı (Şekil 1'deki Ch1, Ch2, Ch3 ve Ch4) için kartezyen koordinat bilgileridir. $S1_{ağırlık}[n]$, $S2_{ağırlık}[n]$ ise n'inci ses kanalının sağ ve sol kulak üzerindeki ağırlıklarını ifade eder.

Hesaplanan ağırlıklar, 0 ile 1 arasında değer alacağından normalize etmeye gerek kalmaksızın, 4 ses kanalının ayrı ayrı kullanıcı kulaklığındaki 2 çalma kanalı üzerindeki ses kazancı olarak kullanılabilir. Örneğin, $S1_{ağırlık}[3] = 0.5$ ise 3'üncü ses kanalının ses seviyesi sağ kulakta yarıya (3dB) düşecektir.

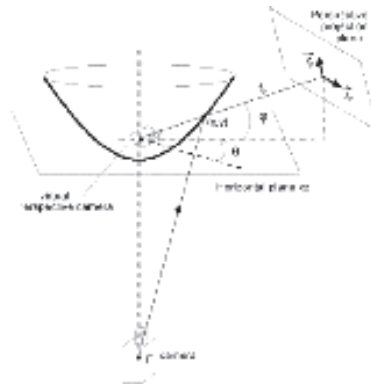
Uzaktan bulunma algısının yaratılabilmesi dört farklı yön için kaydedilen ses verisinin kulaklıkla çalınmak üzere dinamik olarak birleştirilerek iki kanala aktarılması ile mümkün olabilir.

2.2. Görsel Uzaktan Bulunma Sistemi

Çalışmanın görsel kısmında fikir olarak önceden kaydedilmiş tümyönlü videonun bir parçasının alınarak dikdörtgen bir çerçeveye oturtulması için dönüşümü ele alınmaktadır. Bu dönüşüm, tümyönlü video verisinden sadece istenilen bir yöndeki görüntünün elde edilip izlenmesini sağlamaktadır. Bu sistemdeki anahtar nokta, izleme yönündeki görüntünün kullanıcının baş hareketleri ile uyumlu bir şekilde gerçek zamanlı işlenmesi ve elde edilmesidir.

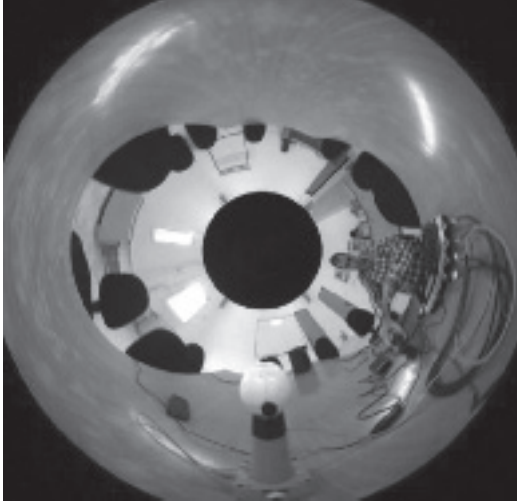
2.2.1. Tümyönlü Video Kaydı

Tümyönlü video yatayda 360 derecelik bir görüş açısı sağlar. Bu video Şekil 4'te görüldüğü üzere tipik olarak konkav aynalı bir düzenekteki yüksek çözünürlüklü kamera ile kayıt edilebilir.



Şekil 4: Tümyönlü kamera modeli(10).

Projenin demo uygulamasında konferans odası uzak ortam olarak seçildi ve tümyönlü video kaydı bu ortamda alındı. 792x792 boyutlarındaki örnek resim çerçevesi Şekil 5'te görülmektedir.



Şekil 5: Tümyönlü videodan örnek resim çerçevesi (792x792).

2.2.2. Radyal-Kartezyen Çevrimi

Tümyönlü video pikselleri ilk yaklaşım ile radyal koordinatlarda varsayılabilir. Tümyönlü videonun video gözlüğünde gösterilmek üzere oluşturulacak dikdörtgen resim çerçevesine dönüşümü için radyal koordinatlardan kartezyen koordinatlara çevirim gerekmektedir.

$$X=r \cos \theta \quad (11)$$

$$Y=r \sin \theta \quad (12)$$

Tümyönlü video çerçevesinde radyal koordinatın merkez noktası çerçevenin orta noktası olarak varsayıldığında (r, θ) radyal parametreleri sırasıyla yarıçap ve X eksenine pozitif yönde yapılan açıyı temsil etmektedir. Kartezyen koordinatlar olarak ifade ettiğimiz (X,Y) parametreleri ise matris formundaki dikdörtgen resim çerçevesinin sırasıyla sütun ve satırlarını temsil etmektedir.

Radyal koordinatlardan kartezyen koordinatlara dönüşüm sonrası merkez kaydırma işlemi gerekmektedir. Bu işlem (13) ve (14)'teki denklemlerde kartezyen parametrelere 396 (Şekil 5'te gösterilen imge için) eklenmesi sonucunda gerçekleştirilir.

$$X=r \cos \theta +396 \quad (13)$$

$$Y=r \sin \theta +396 \quad (14)$$

2.2.3. Çift Doğrusal Interpolasyon Yöntemi ile Geriye Dönük Eşleştirme

792x792 boyutlarındaki tümyönlü video çerçevesinden çıkarılacak ve 640x480 boyutlarındaki dikdörtgen çerçeveye resim kalitesini görece çok bozmadan aktarılacak verinin oluşturulması için çift doğrusal interpolasyon yöntemi ile geriye dönük eşleştirme [11] kullanılmıştır. Bu yöntem, oluşturulacak yeni çerçevenin her piksel değerinin, tümyönlü video çerçevesindeki (13) ve (14)'teki denklemler sonucunda eşleştiği ara değerli noktanın çift doğrusal interpolasyon tekniği ile hesaplanmasına dayanır. Bu yöntem ile ileriye dönük eşleştirme sırasında yaşanabilecek değer atanmamış boşluk oluşması ve aynı nokta için değer atanmasının tekrarlanması gibi problemler engellenmiş olur.



Şekil 6: Çift doğrusal interpolasyon yöntemi ile geriye dönük eşleştirme sonucu elde edilen örnek resim çerçevesi (640x480).

Şekil 6'da görülen, video gözlüğüne aktarılacak örnek resim çerçevesinde, görüş açısı olarak yatayda 75 derece dikeyde ise 200 piksellik boyutlar seçilmiştir.

2.2.4. Hareket Takibi

Kullanıcının baş hareketlerinin takibi uzaktan bulunma sistemlerinin önemli bir noktasıdır. Sensör hassasiyeti ve cevap hızı kullanıcıya verilen hissi oldukça etkilemektedir. Bu projenin demo uygulaması için video gözlüğü üzerindeki 3 serbestlik derecesini algılayan bir sensör kullanılmıştır. Şekil 3'te gösterilen 3 eksenindeki hareketler tümyönlü video çerçevesi üzerinde radyal koordinatlar ile modellenmiştir.

Yaw, θ açısına eklenecek bir parametre, pitch ise r yarıçapına eklenecek bir parametredir. Roll, θ açısına bağlı olarak değişen ve r yarıçapına eklenecek bir parametre olarak modellenmiştir. Sensör bilgileri doğrultusunda dönüşümü yapılacak radyal koordinatlar (r, θ) yenilenmektedir. Böylece 3 serbestlik derecesinde algılanan kullanıcının baş hareketleri takip edilerek kullanıcının baktığı yöndeki video akışı video gözlüğüne aktarılmaktadır.

3. Yazılım

Projenin yazılım kısmında bilgisayar diskinde tutulan tümyönlü video ve 4 kanal ses verileri okutulup ve geçici belleklerde gerekli dönüşüm işlemleri (2. Yöntem) gerçek zamanlı bir şekilde uygulandıktan sonra oynatılmaktadır.

3.1. Kütüphaneler

Temel çoklu ortam oynatma işlemleri için DirectX kütüphaneleri (Ağustos 2009 ve Ekim 2006 sürümleri) kullanılmıştır. İmge ve video işleme alanında oldukça yaygın olan OpenGL, OpenCV gibi kütüphaneler yerine DirectX kütüphanelerinin seçimindeki ana neden ses işleme imkanını da DirectX çatısı altında bulabilmektir. Ayrıca Microsoft'un geniş dökümantasyonu bu kütüphaneyi seçmekte etkili olmuştur. DirectX içindeki DirectSound ve DirectShow bileşenleri, projenin amacına ulaştıracak gerekli metodları içermektedir.

Genel fonksiyonlar ve işlemler için Windows Yazılım Geliştirme Kiti v7.7 kullanılmıştır.

Video gözlüğü (Vuzix VR920 Bundle) üzerindeki sensör verisinin alınabilmesi için üretici firmanın yazılım geliştirme kiti (Vuzix SDK 3.3) kullanılmıştır [12].

3.2. Yazılım Dili ve Geliştirme Ortamı

Yazılım dili olarak C++ seçilmiştir. Gerek yeterliliği gerekse literatürde bir çok örneğinin ve dökümanlarının bulunması bu dili seçmekte etkili olmuştur. DirectX kütüphaneleri yaygın bir şekilde C++ ile kodlanmaktadır.

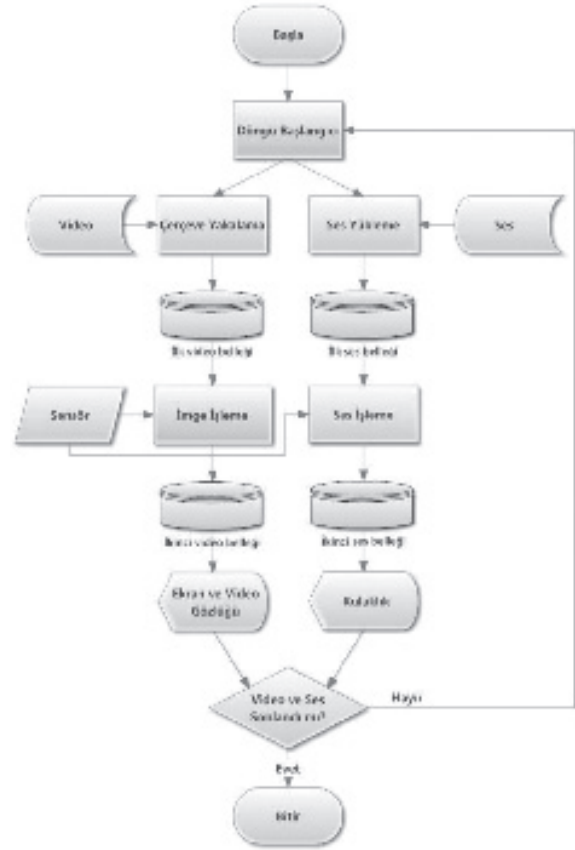
Yazılım geliştirme ortamı olarak DirectX uyumluluğu nedeniyle Microsoft Visual Studio 2010 seçilmiştir.

3.3. Algoritma

Büyük çoğunluğu "2.Yöntem" bölümünde yapılandırılan algoritma kodlanmadan önce ses ve video kısımları için temel çoklu ortam oynatma işlemlerinin uygulandığı örnekler ele alınmıştır.

Ses kısmı için ders olarak anlatılan ve uygulanan, dalga (.wav) formatında bir ses dosyasını yükleme ve oynatma örneği [13] irdelendikten sonra "2.1. İşitsel Uzaktan Bulunma Sistemi" bölümünde modellenen vektörler, hesaplamalar ve metodlar C++ dilinde sınıf olarak kodlanmıştır.

Video kısmında, DirectShow ve MFC (Microsoft Foundation Classes) kullanılarak kodlanan "çerçeve yakalama" tekniği tabanlı bir örnek [14] üzerinde geliştirme yapılmıştır. Gerçek zamanlı bir video işleme tekniği olarak video akışından anlık imge elde edilip imge işleme yöntemi kullanılmıştır. Arka arkaya tekrar edilen bu süreçte videodan imgeye ve yeniden imgeden videoya dönüş sağlanmıştır. Bu döngü Şekil 7'deki akış şemasında görülebilmektedir.



Şekil 7: Akış Şeması

İşitsel ve görsel uzaktan bulunma sistemleri ayrı ayrı geliştirilmiştir. Sınıf olarak kodlanan ses yazılımı, video yazılımı içinde çağırılarak ses nesnesi oluşturulmuştur. Şekil 7'de görüldüğü üzere aynı sensör bilgisini ve işleme döngüsünü paylaşan bu iki sistem, böylece entegre edilmiş olur.

4. Sonuçlar

Sistemin geliştirilmesi sonucunda hedeflenen çıktılarına ulaşılmıştır. Şekil 8'de ekran görüntüsü verilen demo uygulamasında video gözlüğünü ve çift kanallı kulaklığı takıp deneyen insanlar (örnek görsel uygulama Şekil 9'da) işitsel ve görsel uzaktan bulunma hissini bir ölçüde tecrübe etmektedir. Sistemin performansı ve gerçekçiliği hakkında test kullanıcılarından alınan geri bildirimler oldukça olumlu olmakla beraber sistemin geliştirilmesi gereken yönlerini de ortaya koymuştur.



Şekil 8: Demo uygulaması ekran görüntüsü

Yapılan testler sonucunda sistem, kullanıcının baş hareketlerini gecikmesiz bir şekilde algılayarak tepki vermiş ve gerekli işlemleri saniyede 15 çerçeve sayısı (fps) ile kullanıcıya hissettirmeden gerçekleştirmiştir. Görsel kısmın yazılımında gerçekleşen 3 serbestlik derecesindeki hareketlerin işlemleri doğru bir şekilde çalışmaktadır. Kullanıcı uzak ortamdaki odada gerçek zamanlı olarak istediği yere bakabilmektedir. İşitsel kısımda ise doğal işitme hissini gözeterek farklı kaynaklardan gelen seslerin konumunu hissedebilecek düzeyde farkındalık yaratılabilmüş ve sesin görüntü ile tutarlılığı sağlanmıştır.



Şekil 9: Demo Uygulaması

Geribildirimler doğrultusunda geliştirilebilecek bazı konular saptanmıştır. Bunlardan ilki tümyönlü kamera kaydının daha yüksek çözünürlükte alınmasıdır. Her ne kadar 1080P HD seviyesinde kayıt alınmış olsa da videonun küçük bir parçasının video gözlüğüne aktarılması çözünürlükte düşüşe sebep olmaktadır. Ayrıca yazılım ve çalıştırıldığı donanım üzerinde daha etkin bir eniyileme yaparak demo uygulamasında 15 olan saniyedeki çerçeve sayısı artırılabilir. Çözünürlük ve çerçeve hızı iyileştirmelerinin kullanıcıda uyandırılan gerçekçilik hissini arttırması beklenmektedir. Son olarak proje kapsamında olmamasına rağmen “tümyönlü imgenin kalibrasyonu” ele alınabilir. Tümyönlü kameradaki konkav aynanın doğrusal olmayan

karakterinden kaynaklı bozulmayı fiziksel gerçekliğin korunumuyla gidermek için bu kalibrasyon gereklidir.

5. Kaynaklar

- (1) Nagahara, H., Yagi, Y. and Yachida, M., “Wide Field of View Head Mounted Display for Tele-presence with An Omnidirectional Image Sensor”, Graduate School of Engineering Science Osaka University 1-3, Machikaneyama-cho, Toyonaka, Osaka, 560-8531, Japan -2003.
- (2) Morita, S., Yamazawa, K. and Yokoya, N., “Internet Telepresence by Real-Time View-Dependent Image Generation with Omnidirectional Video Camera”, Proc. SPIE 5018, Internet Imaging IV, 51, January 20, 2003; doi:10.1117/12.473592.
- (3) Yamazawa, K., Takemura, H. and Yokoya, N., “Telepresence System with an Omnidirectional HD Camera”, The 5thAsian Conference on Computer Vision, Melbourne, Australia, 23-25 January, 2002.
- (4) Onoe, Y., Yamazawa, K., Takemura, H. and N.Yokoya: “Telepresence by Real-Time View-Dependent Image Generation from Omnidirectional Video Streams”, Computer Vision and Image Understanding, Vol.71, No.2, pp.154-165, 1998.
- (5) Hollerweger, F., “An Introduction to Higher Order Ambisonic”, Order A Journal On The Theory Of Ordered Sets And Its Applications, No. April. (2008) Key, corrected version October 2008
- (6) Fellgett, P., “Ambisonics”, Part One: General system description. Studio Sound, August 1975, pp. 20-40.
- (7) Gonzalez, R. C., and Woods, R. E., “Geometric spatial transformations and image registration” in Digital Image Processing, 3rd ed., Ed. McDonald, M.,2008, pp. 87-88.
- (8) Vuzix Official Website: “Vuzix-Wrap 920 VR Bundle”, Internet:http://www.vuzix.com/consumer/products_wrap920vrbundle.html#vr_manager, (Dec.,2012)
- (9) Hoffman, K. and Kunze, R., “Inner Product Spaces” in Lineer Algebra, 2nd ed., Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1971, pp. 270-277.
- (10) Grassi Junior, V. and Okamoto Junior, J. “Development Of An Omnidirectional Vision System” J. Braz. Soc. Mech. Sci. & Eng. 2006, vol.28, n.1, pp.58-68.
- (11) Gonzalez, R. C., and Woods, R. E., “Image Interpolation” and “Geometric spatial transformations and image registration” in Digital Image Processing, 3rd ed., Ed. McDonald, M.,2008, pp. 65-66, 87-88.
- (12) Vuzix SDK Documentation, “Vuzix Eyewear Software Development Kit Version 3.1”
- (13) DirectX 10 Tutorials: “Tutorial 14: DirectSound”, Internet: <http://www.rastertek.com/dx10tut14.html>, May 17, 2013 (Nov., 2012)
- (14) Zhu, M., “An Easy Video-processing Framework by Grabbing Frames as Bitmaps Using DirectShow”, 2008, Internet: <http://www.codeproject.com/Articles/14689/An-Easy-Video-processing-Framework-by-Grabbing-Fra>, Mar. 5, 2008 (Nov., 2012)