

Ses ve Video İşaretlerinde Saklı Markof Modeli Tabanlı Düşen Kişi Tespiti HMM Based Falling Person Detection Using Both Audio and Video

B.Uğur Töreyn, Yiğithan Dedeoğlu, A.Enis Çetin

Bilkent Üniversitesi, 06800 Ankara
{bugur,yigithan,cetin}@bilkent.edu.tr

Özetçe

Videoda düşen bir kişinin otomatik olarak tespit edilmesi, kapalı devre televizyon ve ev içi destek sistemlerini kapsayan güvenlik alanları için önemli bir uygulamadır. Bu bildiride, video işaretindeki insan hareketi, saklı Markof modelleriyle modellenmiştir. Buna ek olarak, videoyla birlikte kaydedilen ses işaretleri de, düşen bir kişiyle, yere oturan bir kişiyi birbirinden ayırtmada kullanılmıştır. Pek çok video kayıt sistemi, aynı zamanda ses kaydı da yapabilmekte ve düşen bir kişinin çıkardığı ses fazladan bir ipucu olarak bulunmaktadır. Ses işaretine ait sonuçlar da saklı Markof modellerine dayandırılarak elde edilmekte ve video işaretine ait sonuçla birleştirilip son karar verilmektedir.

Abstract

Automatic detection of a falling person in video is an important problem with applications in security and safety areas including supportive home environments and CCTV surveillance systems. Human motion in video is modeled using Hidden Markov Models (HMM) in this paper. In addition, the audio track of the video is also used to distinguish a person simply sitting on a floor from a person stumbling and falling. Most video recording systems have the capability of recording audio as well and the impact sound of a falling person is also available as an additional clue. Audio channel data based decision is also reached using HMMs and fused with results of HMMs modeling the video data to reach a final decision.

1. Giriş

Düşen bir kişinin tespiti, kapalı devre televizyon ve ev içi destek sistemlerini kapsayan güvenlik alanları için önemli bir uygulamadır. Yakın bir gelecekte akıllı evler, içinde bulunan kişilerin yaptıklarını pek çok güvenlik kamerası ve algılayıcılarla izleyip, ona göre işlem yapabilecek ve/veya önlem alabilecek hâle gelecektir. Bu evler, evin içindeki bakıma muhtaç yaşlılar ve çocuklara başlarına kötü birşey geldiğinde, yardım bile çağırabileceklerdir [1,2,3]. Hâlihazırda yaşlılar için tasarlanan ve kullanılan giyilebilir algılayıcılar ivmeölçer, pasif kızılberisi algılayıcı veya basınca duyarlı algılayıcı gibi çeşitli duyucular içermektedir. Ancak, yaşlı ve bakıma muhtaç kişiler bunları giymeyi sıklıkla unutabilmekte ve kimi zaman bu algılayıcılar yanlış alarm verebilmektedir. Yapay görü sistemleri, düşme tespiti için kullanıcıya yük getirmeyen seçenekler sunar. Bu bildiride, video tabanlı bir

düşen kişi tespit yöntemi önerilmektedir. Verilen nihâi kararda video işaretinin hem görüntü hem de ses verileri etkili olmaktadır.

Önerilen video çözümleme algoritması, mevcut imge çerçevesindeki hareketli bölgelerin tespit edilmesiyle başlar. Hareketli bölgeyi çevreleyen en küçük kutu ve buna ait parametreler kestirilir. Böylelikle, videodaki hareketli nesneye ait tek boyutlu bir zaman dizisi sinyali elde edilmiş olur. Bu işaretin dalgacık dönüşümü hesabedilir ve elde edilen yeni işaret, önceden çeşitli insan hareketleriyle eğitilmiş saklı Markof modellerinde kullanılır. Dalgacık alanındaki işaretlerin, zaman alanı sinyallerine göre daha iyi sonuçlar verdiği gözlenmiştir. Bunun nedeni, dalgacık işaretlerinin ilgili sinyaldeki ani değişikliklere duyarlı olması ve işaretin durağan kısımlarını ihmal etmesidir.

Ses çözümleme yöntemi de benzer sebeple dalgacık alanı verisini kullanmaktadır. Kişinin olağan hareketlerine ve düşmesine karşılık gelen saklı Markof modelleri, ses işaretlerine ait kararın verilmesinde temel teşkil etmektedir. Bu karar, video işaretine ait kararla birleştirilip son karar verilmektedir.

Nait-Charif ve McKenna, tüm yönlü bir kameradan elde edilen görüntülerdeki hareketli nesnelerin gezingelerinden düşen kişileri tespit etmişlerdir [4]. Tüm yönlü bir kamera yerine, daha düşük maliyetli bir kamera kullanıldığında hareketli nesnelere ait gezingelerin kestirimi zorlaşmaktadır. Ortamda tüm yönlü bir kamera mevcutsa, önerdiğimiz metodla [4]'te anlatılan yöntem birlikte kullanıldığında çok daha dayanıklı bir sistem elde edilebilir. İnsan hareket gezingesine dayanan bir diğer düşme tespit yöntemi de [5]'te anlatılmaktadır. Ancak, ne [4]'te ne de [5]'te, olayların tespitinde ses işaretleri kullanılmamıştır.

Video çözümleme algoritması ikinci, ses çözümleme algoritması ise üçüncü kısımda anlatılmaktadır. Deneysel sonuçlar dördüncü kısımda sunulmuştur.

2. Video İşaretinin Çözümlemesi

Video işareti çözümleme yöntemimiz üç adımdan oluşmaktadır. İlk adımda videodaki hareketli bölgeler tespit edilir. İkinci adımda hareketli bölgeye ait en-boy oranıyla ilintili işaretin dalgacık dönüşümü katsayıları hesaplanır. Üçüncü ve son adımda ise dalgacık alanı verileri kullanılarak saklı Markof modeli tabanlı sınıflandırma yapılır.

2.1. Hareketli bölge tespiti

Hareketli piksel ve bölgeler [6]'da anlatılan arkaplan kestirimi yöntemiyle bulunmaktadır. Bu yöntem kameranın

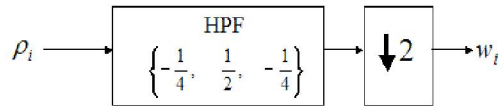
sabit bir konumda bulunduğu varsayımına dayanmaktadır. Kameradan elde edilen görüntüden bir arkaplan resmi oluşturulmaktadır. Mevcut andaki arkaplan resmi, önceki arkaplan resim imgesi ve mevcut andaki resim imgesinden özylenelemeli olarak bir sonsuz dürtü yanıtı süzgeciyle kestirilmektedir. Hareketli bölge tespit yönteminin gerçek zamanda çalışması gerekmektedir. Bu yüzden, kullanılan yöntemin hareketli bölgelerin sınırlarını hassas biçimde tespit edebilmesinden çok, hesaplama karmaşıklığının düşük olması önem kazanmaktadır. Bu açıdan [7] ve [8]'de anlatılan yöntemler yerine arkaplan kestirimine dayalı yöntemi tercih ettik [6].

2.2. Hareketli bölgelerden öznelilik çıkarma

Video çözümlemesindeki ilk adım sonucunda hareketli nesnelere ait en küçük kuşatan kutular bulunur. Daha sonra her nesneye ait en-boy oranı, ρ , hesaplanır. Örneğin, i . hareketli nesneye ait en-boy oranı şu şekilde hesaplanır:

$$\rho_i(n) = \frac{H_i(n)}{W_i(n)} \quad (1)$$

Bu ifadede, $H_i(n)$ ve $W_i(n)$, n . çerçevedeki i . hareketli nesnenin en küçük kapsayan kutusuna ait sırasıyla boy ve en değerleridir. Daha sonra bu en-boy oranlarına ait dalgacık dönüşümü katsayıları Şekil 1'deki yüksek geçiren süzgeç ve alt-örnekleme ikilisinden geçirilerek elde edilir.



Şekil 1: En-boy oranı ρ_i 'ye ait w_i dalgacık dönüşümü katsayıları, tam sayı aritmetiği kullanan yüksek geçiren süzgeçten [9] ve alt-örneklemeden geçirilerek elde edilir.

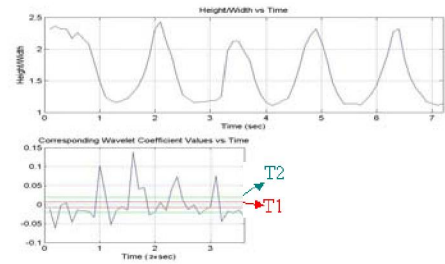
Tek boyutlu en-boy oranına ait dalgacık dönüşümü işareti, saklı Markof modeli tabanlı sınıflandırma için öznelilik işareti olarak kullanılır. Bu öznelilik sinyalinin, yürüyen ve düşen kişiler için farklılık gösterdiği gözlenmiştir. Yürüyen bir kişiye ait en-boy oranında ve bu orana ait dalgacık dönüşümü işaretinde yarı-periyodik bir özellik bulunurken, Şekil 2, düşen veya oturan bir kişiye ait en-boy oranı işaretinde ani bir azalma görülmektedir, Şekil 3. Düşen kişiye ait dalgacık dönüşümü işaretinde de bu düşüş ve sonrasındaki durağan seyir yine Şekil 3'te görülmektedir.

En-boy oranı işareti yerine bu işarete ait dalgacık dönüşümü katsayıları kullanmanın iki önemli getirisi vardır. Bunlardan ilki dalgacık dönüşümünün periyodik olmayan işaretlerdeki bu karakteristiği kolayca ortaya çıkarmasıdır. Bu durum düşen kişi örneğinde çok faydalı olmaktadır. İkinci bir getirisi de, dalgacık dönüşümü işaretleri ortalama değeri sıfır olan sinyaller oldukları için, bu sinyaller için eşik değeri atayabilmek kolaylaşmaktadır. Böylelikle, kameranın görüş alanındaki kişilerin farklı en-boy oranlarına sahip olmalarından kaynaklanan eşik değeri atama zorluğu ortadan kalkmakta, belirlenen eşik değerleri, kişilerin en-boy ölçülerinden bağımsız olmaktadır. Markof modellerindeki çeşitli durumları tanımlamak üzere, Şekil 2'de gösterildiği gibi, dalgacık dönüşümü sinyalleri için $T1$ ve $T2$ olarak iki

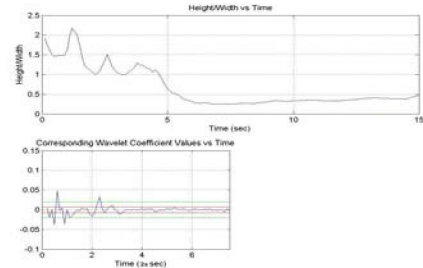
eşik değeri tanımlanmıştır. İlk eşik değeri olan $T1$ dalgacık katsayılarının sıfıra yakınlığını belirtmektedir. Kişinin düşmesinin ardından ideal olarak dalgacık katsayıları sıfır olmalıdır. Ancak gürültü ve düşen kişinin kımıldamaları nedeniyle bu katsayılar sıfır etrafında küçük salınımlar gösterir. Dalgacık dönüşümü katsayılarındaki büyük salınımlar ise $T2$ tarafından yakalanır. Eğer kişiye ait dalgacık dönüşümü katsayıları sıklıkla bu ikinci eşik değeri geçiyorsa, o zaman bu kişi muhtemelen yürümek ya da koşmak gibi periyodik bir hareket sergiliyor demektir.

2.3. Saklı Markof modeli tabanlı sınıflandırma

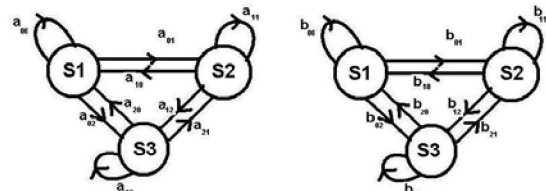
Bu bildiride, bir kişinin hareketinin sınıflandırılması amacıyla biri yürümeye, diğeri ise düşmeye karşılık gelen iki tane üç durumlu Markof zinciri kullanılmıştır, Şekil 4a ve 4b. Sıfırdan büyük $T1 < T2$ eşik değerleri bu saklı Markof modellerinin üç durumunu tanımlamaktadır.



Şekil 2: Yürüyen bir kişinin en-boy oranı ρ 'ya ait zamana göre yarı-periyodik işaret (üstte), ve bu işarete ait dalgacık dönüşümü sinyali (altta). Dalgacık alanında tanımlanan $T1$ ve $T2$ eşik değerleri, kişilerin en-boy oranı farklılıklarına dayanıklıdır.



Şekil 3: Düşen bir kişiye ait tek boyutlu en-boy oranı işaretinin zamanla değişimi (üstte), ve bu değişime ait dalgacık dönüşümü sinyali (altta).



Şekil 4: (a) Yürümeye ve (b) düşmeye karşılık gelen üç durumlu Markof modelleri

Eğer n anında $|w_i(n)| < T1$ ise model S1 durumunda, eğer $T1 < |w_i(n)| < T2$ ise S2 durumunda, ve eğer $|w_i(n)| > T2$ ise S3 durumunda bulunmaktadır. Saklı Markof modelinin eğitim aşamasında geçiş olasılıkları olan a_{uv} and b_{uv} , $u, v = 1, 2, 3$, değerleri, eğitim videoları kullanılarak kestirilir. Deneylerimizde, bu olasılıkların kestirimi için yirmi ardışık video resmi kullanılmıştır.

Yürüyen bir kişinin hareketi yarı-periyodik olduğu için, Markof zincirindeki geçiş olasılıklarının yaklaşık aynı değerlere sahip olmasını bekleriz. Bu yüzden a değerleri birbirine yakındır. Ancak, düşen bir kişinin dalgacık dönüşümü işareti sıfıra yakın değerler almaya başlar. Bu yüzden b_{00} 'ın, diğer tüm olasılık değerlerinden daha yüksek olması beklenir.

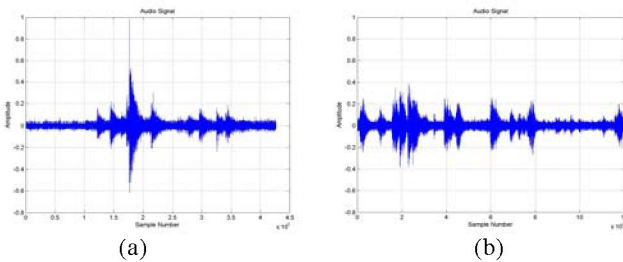
Hareketin tanınması aşamasında, nesneye ait son yirmi en-boy oranı değeri ve bu değerlere ait dalgacık dönüşümü katsayıları hesaplanır. Böylelikle yirmi elemanlı bir durum dizisi elde edilir. Yürümeye ve düşmeye karşılık gelen Markof modellerinden, bu dizi için en yüksek olasılık değerini üreten modele karar verilir. Ancak yalnızca video çözümü, bir kişinin düştüğünün anlaşılması için yeterli değildir. Yere yavaşça oturan bir kişi de düşmeye benzer bir durum dizisi oluşturabilir. Bu yüzden ses verisinin de çözümü gerekmektedir.

3. Ses İşaretinin Çözülmesi

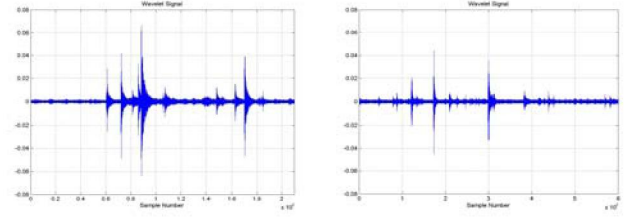
Yazarların konuşma işleme konusundaki önceki çalışmalarında, dalgacık alanında çıkarılan öznelitektörlerinin, Fourier alanından elde edilen öznelitektörlerine göre daha dayanıklı olduğu sonucuna varılmıştır [10]. Bu yüzden, video sinyali çözümlemesine benzer şekilde, ses işareti çözümlemesi yöntemi de dalgacık alanındaki analize dayandırılmaktadır. Ses çözümleme yöntemimiz üç aşamadan meydana gelmektedir. İlk aşamada dalgacık dönüşümü sinyali elde edilir, daha sonra bu işaretten öznelitektör çıkarımı yapılır ve son olarak da saklı Markof modeli tabanlı sınıflandırma yapılır.

3.1. Dalgacık işareti

Ses işaretine ait dalgacık dönüşümü katsayılarının elde edilmesinde Şekil 1'deki sistem kullanılmaktadır. Şekil 5a ve 5b'ye ait dalgacık sinyalleri Şekil 6'da sunulmuştur.



Şekil 5: (a) Örnek numarası $1,8 \times 10^5$ etrafında meydana gelen düşmeye ait ses sinyali, ve (b) konuşma ($0-4,8 \times 10^5$), eğilme ($4,8 \times 10^5-5,8 \times 10^5$), konuşma ($5,8 \times 10^5-8,9 \times 10^5$), yürüme ($8,9 \times 10^5-10,1 \times 10^5$), eğilme ($10,1 \times 10^5-11 \times 10^5$), ve konuşma ($11 \times 10^5-12 \times 10^5$) durumlarına karşılık gelen ses sinyalleri.



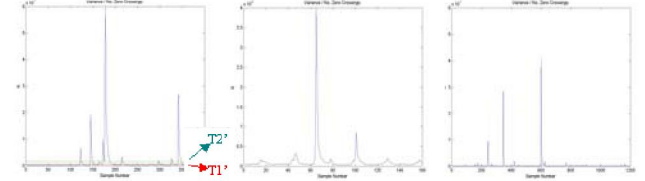
Şekil 6: Düşme ($0,9 \times 10^5$) (solda), ve konuşma ($0-2,4 \times 10^5$), eğilme ($2,4 \times 10^5-2,9 \times 10^5$), konuşma ($2,9 \times 10^5-4,5 \times 10^5$), yürüme ($4,5 \times 10^5-5 \times 10^5$), eğilme ($5 \times 10^5-5,5 \times 10^5$), ve konuşma'ya ($5,5 \times 10^5-6 \times 10^5$) karşılık gelen dalgacık işaretleri, (sağda).

3.2. Dalgacık işaretlerinin analizi

Ses sinyallerine ait dalgacık işaretleri, beşyüzer örneklik pencerelerde analiz edilmiştir. Kullanılan ses sinyalinin örnekleme sıklığı 44,1 kHz'tir. Her i . pencereye ait değişinti (varyans) σ_i^2 ve sıfır seviyesi kesme sayısı, Z_i , belirlenmiştir. Yürümeye ait ses sinyalinin değişinti ve sıfır kesme sayıları yarı-periyodik bir karakteristiğe sahiptir. Buna karşılık sendeleyip düşen birine ait ses işaretinin sıfır kesme sayısı azalırken, değişinti değerinde bir artış gözlenmiştir. Biz de bu gözlemlerimize dayanarak her bir pencere için κ olarak adlandırdığımız bir parametre tanımladık:

$$\kappa = \frac{\sigma_i^2}{Z_i} \quad (2)$$

κ parametresi daima pozitif değerler almaktadır.

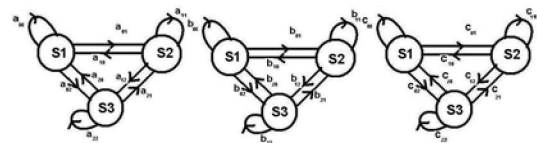


Şekil 7: Düşmeye (180) (solda), yürümeye (ortada), ve konuşmaya (0-480), eğilmeye (480-590), konuşmaya (590-900), yürümeye (900-1000), eğilmeye (1000-1100), ve konuşmaya (1100-1200) ait κ parametresi değişimi (sağda). $T1'$ ve $T2'$ eşikleri, κ alanında tanımlanmıştır.

Konuşma sırasında söylenen sesli harflere ait κ parametresi sessiz harflere ait parametre değerinden daha büyüktür. Farklı durumlar için gözlenen κ parametresinin örnek numarasına göre değişimi Şekil 7'de gösterilmiştir.

3.3. Saklı Markof modeli tabanlı sınıflandırma

Ses işaretlerinin sınıflandırılması amacıyla, video işaretlerine benzer şekilde, Şekil 8'de gösterilen, yürümeye, konuşmaya ve düşmeye karşılık gelen üç durumlu üç Markof modeli kullanılmaktadır.



Şekil 8: Yürüme(solda), konuşma(ortada) ve düşme (sağda) seslerinin sınıflandırılması için kullanılan üç durumlu üç Markof modeli

Eğer i . penceredeki dalgacık sinyaline ait κ parametresi için $\kappa_i < T1$ doğruysa model S1 durumunda, eğer $T1 < \kappa_i < T2$ doğruysa S2 durumunda ve eğer $\kappa_i > T2$ doğruysa S3 durumundadır. Modellere ait geçiş olasılıkları, yirmi ardışık dalgacık işareti penceresi kullanılarak kestirilmektedir. Seslerin sınıflandırılması sırasında, ardışık yirmi κ parametresi kestirilmekte ve bu parametrelere karşılık gelen durum dizisi çıkarılmaktadır. Bu durum dizisi Şekil 8'deki üç modele beslenerek en yüksek olasılığı üreten model tespit edilmektedir. Ses verisine ait bu sonuçla, video işareti sonucu "mantıksal ve" işlemiyle birleştirilmekte ve son karar bu şekilde verilmektedir.

4. Deneysel Sonuçlar

Önerilen yöntem AMD AthlonXP 2000+ 1.66GHz işlemcili bir bilgisayarda kodlanmış ve gerçek zamanda çalışmaktadır. Testlerde, toplam 15.823 resimli 64 video kullanılmıştır. Tüm videolarda kamera sabit olup görüş alanında her an yalnızca tek bir kişi hareket etmektedir. Videolara ait içerik bilgisi Tablo 1'de özetlenmiştir.

Tablo 1: Test kümesi içerik özeti

Video içeriği	Ses işareti var mı?	Video sayısı
Yürüme + Konuşma	Evet	16
Oturma + Konuşma	Evet	5
Oturma	Evet	4
Yürüme + Düşme	Evet	25
Yürüme + Düşme	Hayır	14

Tablo 1'deki kimi videolara ait örnek resimler Şekil 9'da sunulmuştur. Bu test kümesine ait sonuçlar ise Tablo 2'de gösterilmiştir. Yalnızca video işareti kullanılarak, isteyerek oturan biriyle düşen birisini ayırtmak mümkün değildir. Her iki işaret kullanıldığında ise, Tablo 2'de görüldüğü gibi, oturan bir kişi için düşme kararı verilmemektedir.

Tablo 2: Test kümesine ait tespit sonuçları

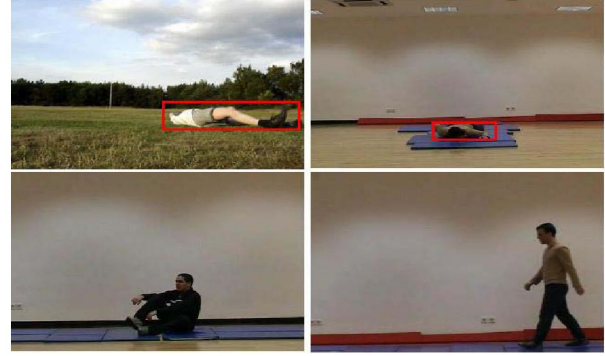
Video içeriği	Ses işareti var mı?	Video sayısı	Düşme tespit edilen video sayısı	
			Video	Ses+Video
Yürüme +Konuşma	Evet	16	0	0
Oturma+ Konuşma	Evet	5	5	0
Oturma	Evet	4	4	0
Yürüme + Düşme	Evet	25	25	25
Yürüme + Düşme	Hayır	14	14	14

5. Sonuç

Videoda düşen bir kişinin otomatik olarak tespiti için bir yöntem geliştirilmiştir. Bu çalışmanın en önemli katkısı düşme kararını hem video hem de ses işareti çözümlemeleri sonucunda vermesidir. Ses bilgisi, oturan bir kişiyle düşen bir kişinin ayırdedilmesi açısından gereklidir. Olayların sınıflandırılmasında üç durumlu Markof modelleri kullanılmıştır. Markof model öznelikleri, video nesnelerini kapsayan kutulara ait zamansal tek boyutlu sinyallerin dalgacık katsayılarından elde edilmiştir. Dalgacık sinyalleri sıfır ortalamalı oldukları için Markof durumlarının

tanımlanması kolaylaşmış, böylelikle nesne boyutlarındaki farklılıklara daha dayanıklı bir yöntem elde edilmiştir.

Önerilen yöntemin hesaplama karmaşıklığı düşük olduğu için kişisel bilgisayarlarda gerçek zamanda çalışabilmektedir. Benzer saklı Markof model yapıları, aniden duran araçların veya otoyallardaki kazaların otomatik tespitinde de kullanılabilir.



Şekil 9: Düşme, oturma ve yürüme videolarından örnek resimler

6. Kaynakça

- [1] Barnes, N.M., Edwards, N.H., Rose, D.A.D., Garner, P.: Lifestyle Monitoring: Technology for Supported Independence. IEE Comp. and Control Eng. J. (1998).
- [2] Bonner, S.: Assisted Interactive Dwelling House: Edinvar Housing Assoc. Smart Tech. Demonstrator and Evaluation Site In: Improving the Quality of Life for the European Citizen (TIDE), (1997) 396–400.
- [3] McKenna, S.J., Marquis-Faulkes, F., Gregor, P., Newell, A.F.: Scenario-based Drama as a Tool for Investigating User Requirements with Application to Home Monitoring for Elderly People. In Proc. of HCI, (2003).
- [4] Nait-Charif, H., McKenna, S.: Activity Summarisation and Fall Detection in a Supportive Home Environment. In Proc. of ICPR'04, (2004) 323-326.
- [5] Cuntoor, N.P., Yegnanarayana, B., Chellappa, R.: Interpretation of State Sequences in HMM for Activity Representation. In Proc. of IEEE ICASSP'05, (2005) .
- [6] Collins, R.T., Lipton, A.J., Kanade, T., Fujiyoshi, H., Duggins, D., Tsin, Y., Tolliver, D., Enomoto, N., Hasegawa, O., Burt, P., Wixson, L.: A System for Video Surveillance and Monitoring: VSAM Final Report. Tech. Report CMU-RI-TR-00- 12, CMU (1998).
- [7] Bagci, M., Yardimci, Y., Cetin, A.E.: Moving Object Detection Using Adaptive Subband Decomposition and Fractional Lower Order Statistics in Video Sequences. Elsevier, Signal Processing. (2002) 1941—1947.
- [8] Stauffer, C., Grimson, W.E.L.: Adaptive Background Mixture Models for Real-Time Tracking. Proceedings of IEEE CVPR'99 246-252, 1999.
- [9] Kim, C.W., Ansari, R., Cetin, A.E.: A class of linear-phase regular biorthogonal wavelets. In Proc. of IEEE ICASSP'92 (1992) 673-676.
- [10] Jabloun, F., Cetin, A.E., Erzincan, E.: Teager Energy Based Feature Parameters for Speech Recognition in Car Noise. IEEE Signal Processing Letters (1999) 259-261.