

İşaret Dili Hareketlerinin İzlenmesi ve İki Boyutlu Özniteliklerden Üç Boyutlu Hareket Sentezlenmesi

Sign Language Motion Tracking and Generating 3D Motion Pieces Using 2D Features

Hamdi Dibeklioglu*, Erinç Dikici**, Pınar Santemiz*, Koray Balci*, Lale Akarun*

* Bilgisayar Mühendisliği Bölümü

** Elektrik ve Elektronik Mühendisliği Bölümü

Boğaziçi Üniversitesi, İstanbul

{hamdi.dibeklioglu, erinc.dikici, pinar.santemiz, balci, akarun}@boun.edu.tr

Özetçe

Bu çalışmada, stereo kameradan oluşan bir kurulumda üç boyutlu hareketi izleyen ve yeni işaret parçaları sentezleyen bir sistem tasarlanmıştır. Bu sistemde işaretleyici kullanarak oluşturulmuş bir veri kümesi kullanılmış ve işitme engelliler Türk işaret dilinden ifadeler canlandırılmıştır. Hem işaretleyicilerin merkezlerinin bulunması sırasında hem de bulunan noktaları izleme sırasında ilişkilik problemi çözülmeye çalışılmıştır. Nokta bulma ve hareket izleme için önerilen yeni yöntemlerin başarımları karşılaştırılmıştır. Takip edilen ve boyutları düzgelenecek gezinimler Saklı Markov Modelleri (SMM) ile sınıflandırılmış, bu modeller kullanılarak yeni işaret parçaları oluşturulmuştur.

Abstract

In this study we designed a system that can track 3D motion and generate new pieces of motion, using a stereo camera setup. A dataset from some expressions of Turkish sign language is built using markers. In order to find the center of the markers and track their coordinates the correspondance problem had to be solved. Results of newly proposed marker detection and tracking algorithms are compared. Finally, following the scale normalization, the trajectories are modelled using Hidden Markov Models (HMM). Using these models, new signs are generated.

1. Giriş

İşaret dili, işitme engellilerin birbirleriyle ve yaşadıkları toplumdaki kişilerle anlaşmasını sağlayan en önemli iletişim yöntemlerinden biridir. Son dönemde, bilgisayar ve kamera teknolojisindeki yenilikler sayesinde işaret dilinin kullanımı ve öğrenilmesini kolaylaştıracak yöntemlerin de önü açılmış, hareket izleyen, tanıyan ve canlandıran yöntemler bu amaçla kullanılmaya başlanmıştır.

İşaretleyici yardımıyla hareketi izleyen sistemlerde, nokta bulma ve izleme yöntemleri kullanılarak gezinimler çıkarılması amaçlanır. Gezinimin çıkarılmasında çeşitli hareket izleme teknikleri kullanılabileceği gibi (Hough Dönüşümü [1], Kalman Süzgeci [2]), zamansal şablonlara dayalı teknikler de ilişkilik problemini çözmedeki başarıları nedeniyle kullanılabilir [3-5].

Çıkarılan gezinimler kullanılarak hareket dizilerinin ait oldukları sınıfı modellemek ve yeni işaret parçaları sentezlemek mümkündür. Bunun için Saklı Markov Modelleri (SMM) kullanan yöntemler bulunmaktadır [6,7].

Bu çalışmada stereo kameradan oluşan bir kurulumda işaret diline ait ifadeler işaretleyici yardımıyla izlenmiş ve SMM kullanılarak yeni işaret parçaları üretilmiştir. 2. kısımda oluşturulan veri kümesi ile ilgili bilgi verilmekte, 3. kısımda önışleme ve işaretleyici bulma yöntemi anlatılmakta daha sonra 4. kısımda işaretleyicilerin koordinat bilgisinin çıkarılması yönteminin detayları verilmektedir. 5. kısımda bu noktaların izlenmesi için kullanılan iki yöntem sunulmakta, son olarak da 6. kısımda SMM uygulaması anlatılmaktadır. Çalışmanın genel sonucuna, izleme algoritmalarının sonuçlarına ve karşılaştırılmasına ise 7. kısımda değinilmiştir.

2. Görüntü Veri Kümesi

Çalışma için 8 kişinin hareketlerini içeren bir veri kümesi oluşturulmuştur. Bu veri kümesinde, işitme engelliler Türk İşaret Dili'ne (TİD) ait “merhaba”, “hoşçakal”, “evet”, “hayır” ve “teşekkürler” ifadeleri canlandırılmaktadır. Kişilerin omuz, dirsek, el, alın ve karın bölgelerine toplamda 8 adet sarı renkte işaretleyici cisim yerleştirilmiştir (Şekil 1).



Şekil 1: Veri kümesinden örnek görüntüler.

Görüntüler Point Grey Bumblebee® stereo kamera sistemi kullanılarak 640x480 boyutunda ve 30 fps hızında, ortalama 5 sn süreli videolar halinde kaydedilmiş ve taşınabilir imgecik

eşleştirme dosya formatına (Portable Pixelmap - PPM) dönüştürülmüştür. Her bir kişiden “evet” ve “teşekkürler” ifadeleri için beşer ve diğerlerinden birer çift (sol-sağ) olmak üzere toplamda 14 çift görüntü toplanmıştır. Hazırlanan veri kümesi yaklaşık 33 bin çerçeveden oluşmakta ve toplamda 10 GB’lık yer kaplamaktadır.

3. Önışleme ve İşaretleyici Bulma

Toplanan veri kümesindeki görüntülerin (imgelerin) önışlemesi, çalışmanın ilk adımını oluşturmaktadır. Bu işlemin temel amacı, görüntüdeki bilgi taşımayan kısımların elenerek asıl ilgi alanı olan işaretleyici pozisyonlarının belirlenmesidir. Bu amaçla renk bilgisinden faydalanılması düşünülmüş ve işaretleyicilerin (sarı) rengini gürbüz biçimde tanımlayacak iki farklı yöntem uygulanmıştır:

- **Yöntem 1:**

İşaretleyicileri belirleyebilmek amacıyla öncelikle her bir görüntü çerçevesinden, arka plana ait kısımlar imgecik (piksel) karşılaştırma işlemi ile çıkarılır. Arka plan çıkarma işlemi, duvardaki (ışıklandırma sonucu oluşan) sarı tonların yok edilmesi için kullanılmaktadır. Ardından sarı tonların ayırt edilebilmesi amacıyla çeşitli renk uzaylarında eşikleme yapılarak elde edilen görüntünün mavi bandı üzerinde ikilileştirme uygulanır. Sonuç ikili görüntüsündeki bağlantılı bileşenler bulunur ve belirli bir boyuttan küçük bileşenler silinerek olası gürültü elemanlarının yok edilmesi sağlanır.

- **Yöntem 2:**

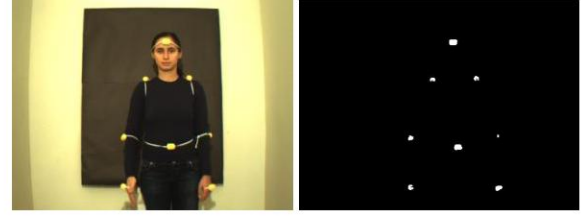
İlk yöntem ile, işaretleyicilerin pozisyonlarının doğru belirlendiği, ancak tespit edilen işaretleyici sayısının beklenen değerler ile uyumsuz olduğu görülmüştür. Bu durumun, hareket sırasında oluşan gölgelerin sarı tonları değiştirmesinden ve gürültü ayıklamada seçilen boyut bilgisinin eşik seviye niteliğine uygun olmamasından kaynaklandığı anlaşılmıştır. Bu nedenle söz konusu olumsuzlukları ortadan kaldıracak farklı bir renk uzayına ihtiyaç duyulmuştur.

Uygulanan ikinci yöntemde sarı renkli işaretleyicilerin diğer tonlardan ayrıştırılması amacıyla YES uzayı seçilmiştir [8]. Bu renk uzayının renklilik (chrominance) bileşeni olan S bandı sarı rengi, insan teninden ve beyaz tonlardan ayırarak, ortam aydınlatmasından bağımsız ve gürbüz biçimde belirlemektedir. RGB uzayı ile YES uzayı arasındaki dönüşüm,

$$\begin{bmatrix} Y \\ E \\ S \end{bmatrix} = \begin{bmatrix} 0.253 & 0.684 & 0.063 \\ 0.500 & -0.500 & 0.000 \\ 0.250 & 0.250 & -0.500 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$

ifadesi ile tanımlanır. İncelenen veri kümesinde S bandındaki değeri belirli bir eşik üzerinde bulunan imgecikler sarı renkli işaretleyici imgecikleri olarak tanımlanır ve oluşturulan ikili imgede beyaz ile gösterilir.

Oluşturulan ikili imgede gürültü olarak tanımlanabilecek istenmeyen imgecikler de görülebilmektedir. Bu olumsuz etkileri yok etmek amacıyla, ikili görüntüye bir alan doldurma (flood filling) işlemi uygulanarak sınırları bir çevre oluşturacak biçimde bağlantılı olan yapıların içleri doldurulur. Son olarak, 3x3’lük yapı elemanı ile morfolojik açma işlemi kullanılarak gürültünün tamamen yok edilmesi sağlanır. S bandında çift eşikleme uygulanmasının da gürültü imgeciklerinin elenmesinde önemli bir rol oynadığı görülmüştür. Veri kümesinden alınan bir örnek görüntü ile, işlemler yapıldıktan sonra elde edilen işaretleyici konumları Şekil 2’de sunulmaktadır.



Şekil 2: İşaretleyici pozisyonlarının elde edilmesi.

4. İşaretleyici Konumlarının Tespiti

İşaretleyicilerin bulunmasından sonra sıra koordinat bilgisinin çıkarılmasına gelmektedir. Her bir işaretleyici merkezi için bir (x,y) koordinat ikilisi elde etmek amacıyla önce literatürde kullanılan Hough çember bulma algoritması incelenmiştir [4]. Ancak bu algoritmanın mevcut yapı ile yeterince başarılı çalışmadığı görülmüştür. Özellikle işaretleyicilerin hareket sırasında örtülmesi sonucu, görüntüleri çemberden çokça farklılaşmakta ve bu durum, algoritmanın işaretleyicileri tespit etmesini zorlaştırmaktadır. Bu nedenle söz konusu durumlarda gürbüz biçimde çalışacak yeni bir yöntem önerilmektedir.

Bu yöntemde, işaretleyicilerin bulunduğu alanları tespit edebilmek için önce imgedeki bağlantılı bileşenler bulunur. Her bir bağlantılı bileşen, bir işaretleyiciyi temsil etmektedir ve bu bileşene ait imgeciklerin (x,y) koordinat ortalamaları, işaretleyicinin merkez konumunu belirtmektedir.

İdeal durumda tüm işaretleyiciler birbirlerinden belli bir uzaklıkta konumlanmaktadır. Bu nedenle başlangıçta tüm gruplar birbirlerinden en az 20 imgecik uzaklıkta olacak şekilde kaynaştırılır. İki koordinatın birbirine bu eşik değerinin altında kalacak kadar yakınlaşması durumunda ise, işaretleyicilerin bir önceki konumlarıyla karşılaştırma yapılarak birbirine yakın noktalar için yeniden bir kaynaştırma işlemi uygulanır. Yine bir önceki konumlarla karşılaştırma yapılarak, gürültü nedeniyle fazladan bulunmuş koordinatlar elenir.

5. İşaretleyicilerin İzlenmesi

Koordinatları tespit edilen noktaları izlerken, işaretleyicinin bir önceki pozisyonuna yakın bir yere gideceği ve yönünü fazla değiştiremeyeceği varsayılır [5]. İzleme için geliştirilen yöntemlerin ilkinde, her adımda daha önceki iki çerçevedeki koordinat bilgileri kullanılarak noktaların bir önceki çerçevedeki pozisyonları ve yönleri bulunur, bunlarla karşılaştırma yapılır ve yeni bulunan noktalar isimlendirilir. Bu nedenle ilk çerçevenin analitik koordinatlarının sıralı ve eksiksiz bir şekilde sisteme verilmesi gereklidir. İşaretleyicinin kaybolması durumunda ise isimlendirilemeyen noktaya 0 değeri atanır. İzleme işlemi bittikten sonra bu kısımlar, noktanın kaybedildiği andan bulunduğu ana kadar doğrusal bir gezipgesi olduğu varsayılarak düzeltilir.

Bu yöntem, izlenmek istenen noktalardan birkaçı birbirine çok yaklaştığı ya da üstüste geldiği durumlarda başarısız olabilmekte ve bu durum bulunan noktaların kaybına yol açabilmektedir. Örtmeye maruz kalmayan noktaların bulunma başarısı hayli yüksek olmakla birlikte, örtmenin yaşandığı andan itibaren, çakışmaya maruz kalan noktaların bulunma başarısı düşmektedir.

İkinci yöntemde, ilk iki çerçevenin analitik koordinatları sıralı ve eksiksiz kabul edilerek hesaplamalar yapılmaktadır. Bu yöntemde önce işaretleyiciler sıralanır, sonra da eksik işaretleyiciler kestirilir. Her çerçeve Fr_k ve o çerçevedeki takip edilen işaretleyiciler P_{ki} olarak kabul edilirse, k numaralı çerçeve nokta sayısı N_{k_p} her çerçevede farklılık gösterebilir.

Eğer ilgili çerçevedeki işaretleyici sayısında eksiklik varsa, bu bilgiler, önceki iki çerçeve ve daha karmaşık durumlarda sonraki çerçeveler kullanılarak doğrusal kestirilir.

5.1. İşaretleyicilerin Sıralanması

İşaretleyicilerin birbirinin üstünden geçmesi nedeniyle, etiketlemede yer değiştirme olabilmektedir. Bu problemi çözmek için, Döngülü En Yakın Nokta (Iterative Closest Point - ICP) Algoritması kullanılarak, üçüncü çerçeveden itibaren, yeni çerçevedeki işaretleyici koordinatları ile önceki çerçevedeki koordinatların yer değiştirme matrisi hesaplanır. İki çerçevedeki en yakın işaretleyiciler aynı işaretleyici olarak belirlenir. Bu işlem her basamakta üçer çerçeve kullanılarak döngüsel olarak tüm görüntü dizisine uygulanır. Yer değiştirme matrisi, işaretleyicilerin iki çerçeve arasındaki yer değişiminin Öklid uzaklığı cinsinden hesaplanmasıyla oluşturulur. Sistem yer değiştirme matrisini kullanarak, yeni çerçevedeki işaretleyicileri bir önceki sıralamaya uygun olacak şekilde tekrar sıralar.

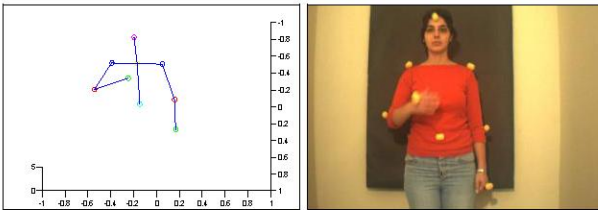
5.2. Eksik İşaretleyicilerin Tahmin Edilmesi

5.2.1. İleriye Dönük Kestirim Algoritması: İşaretleyicileri sıralama işlemi sırasında eksik işaretleyiciler bulunur ve oluşturulan işaretleyici dizisinde -1 olarak gösterilir. Eksik işaretleyiciler bulunduktan sonra gerekli koordinatlar basit bir kestirim algoritması ile hesaplanır. Kestirim işleminde yer değiştirme katsayıları, ilgili çerçeveden önceki iki çerçeve arasında işaretleyicilerin yer değişiminin Öklid uzaklığı cinsinden hesaplanmasıyla bulunur. İşlenen çerçevedeki eksik işaretleyici koordinatları bir önceki çerçevedeki ilgili analitik koordinatlara yer değiştirme katsayıları eklenerek bulunur. Bu yöntem döngüsel olarak kullanılarak tüm eksik işaretleyici koordinatları kestirilir.

5.2.2. İki Yönlü Doğrusal Aradeğerleme: Birden fazla sıralı çerçevede, aynı işaretleyici eksik olduğu durumlarda doğruluk oranını artırabilmek için ilave bir kestirim algoritmasına gerek duyulmaktadır. İlave kestirim işleminde, ardarda gelen eksik işaretleyici bilgileri ve ilgili çerçevelerin sıra numaraları yığın hafıza bölgesinde (stack) depolanır. Eksik işaretleyiciler, doğrusal aradeğerleme ile kestirilir.

5.3. Sistem Çıktıları

Tamamlanmış ve sıralanmış işaretleyici koordinatları dizisi ve işaretleyicilerin anlık yer değişimlerini gösteren animasyon görüntüsü sistem çıktısı olarak verilir.



Şekil 3: İşaretleyicilerin çıktı olarak gösterimi.

İki boyutlu imgelerde izleme yapıldıktan sonra, sağ ve sol imgeler ve kameranın kalibrasyon matris bilgisi kullanılarak üç boyutlu koordinatlar elde edilmektedir. Bu aşamada önce sol resimde bulunan her imgeciğin, sağ resimde bu imgeciğe karşı düşen imgeciğin bulunur, yatay yer değiştirme tespit edilir ve böylece aykırılık matrisi hesaplanır. Kullanılan kameranın kalibrasyon parametreleri kesin olarak belirlenebildiği için, derinlik bilgisi sadece bu aykırılık matrisi ve parametreler

kullanılarak bulunur ve üç boyutlu koordinatlar elde edilir. Ek olarak takip edilen işaretleyicilerin yerlerinin anlık değişimleri, basit bir animasyonla gösterilir.

6. SMM Uygulaması

Daha önceki aşamalarda oluşturulan aynı ya da farklı uzunluktaki 2 ya da 3 boyutlu koordinat dizileri kullanılarak, bu dizilerin ait oldukları sınıfın gizli parametreleri bulunur ve her sınıf için farklı bir Saklı Markov Modeli oluşturulur. Daha sonra bu model kullanılarak, o sınıfa dahil olacak yeni hareket parçaları modellenir. Daha sonra elde edilen dizi, anlık yer değişimlerini gösteren bir animasyon görüntüsü ile görüntülenir.

Bu uygulamada, her hareket için 8 ayrı kişiden toplanmış 40 farklı veri kullanılır. Veri toplanan kişilerin fiziksel özellikleri farklı olabileceğinden, algoritmanın düzgün çalışabilmesi için, işaretleyicilerin ölçeklendirilmesi gerekmektedir. Koordinat dizisi, bu ölçeklendirme doğrultusunda değiştirilir. Ayrıca çekim esnasında yarım kalmış ya da düzgün başlamamış, hatalı hareketler SMM'in eğitiminde kullanılmazlar.

Çalışmada en elverişli değişkenleri saptamak amacıyla, algoritma farklı değişkenler kullanılarak çağırılmıştır. Bunun sonucu olarak, durumlar arası geçiş olasılığı küçük alındığında oluşturulan animasyonda hareketin olmadığı görülmüştür. Bu olasılık gereğinden büyük alındığında ise, normalde olması gerektiğinden farklı hareket elde edildiği saptanmıştır. Sonuç olarak, bu olasılık 0.6 alındığında en iyi sonucun elde edildiği görülmüştür.

Çeşitli durum sayıları verilerek sonucun nasıl değiştiği de gözlenmiştir. Genelde, SMM tanıma için kullanıldığında durum sayısını yüksek tutmak, hareketin ezberlenmesine neden olduğu için pek tercih edilmez. Ancak bu çalışmada hareketi taklit etme amacıyla SMM kullanıldığı için, durum sayısının artırılması daha iyi sonuç vermiştir. Bu nedenle durum sayısı en az 10 alınmıştır.

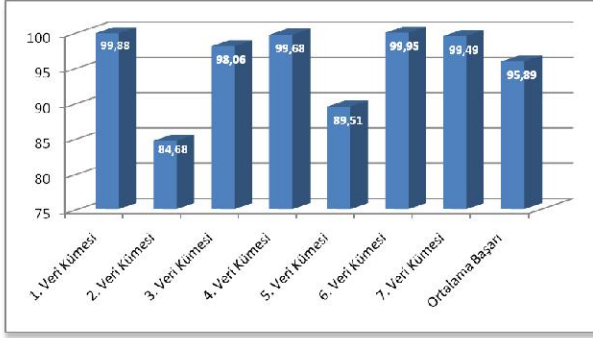
7. Sonuçlar ve Tartışma

Sistem değerlendirmesi 7 kişiden alınmış 5'er "teşekkürler" ifadesi görüntü dizisi ile yapılmıştır. Görüntü dizeleri çift çerçevelerden (sağ-sol) oluşturulmuştur. Toplamda 10346 çerçeve kullanılmıştır.

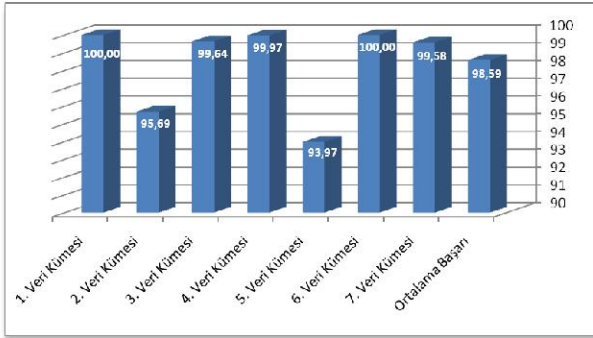
İşaretleyicilerde sarı rengin seçilmesinin çalışmanın zorlayıcı koşullarından biri olduğu görülmüştür. Sarı, gerek insan teni rengine yakınlığı, gerekse uygun olmayan ışıklandırma sonucu oluşan gölgelerden kolayca etkilenebilmesi nedeniyle diğer tonlardan ayırt edilmesi güç olan renklerden biridir. Literatürde sıkça kullanılan renk uzaylarında (RGB, HSI, CMY, vb.) belirtilmiş eşik değer seviyelerinin, çekim ortamındaki aydınlatma koşullarına uyumsuzluğu nedeniyle başarılı sonuç vermediği gözlenmiştir [9,10].

İzleme algoritmalarının değerlendirmesi işleminde işaretleyicilerin sırası da göz önüne alınarak gerçek koordinatlardan en fazla 5 imgeciğin uzaklıkta bulunan işaretleyiciler başarılı kabul edilmiştir. İki algoritma karşılaştırıldığında, örtme olmayan durumlarda sonuçların yakın olduğu görülmüştür. Ancak örtme olan durumlarda birinci izleme algoritmasının yeterince başarılı çalışmadığı görülmüştür.

Şekil 4 ve 5, iki değişik izleme algoritmasında işaretleyici konumlarındaki başarıyı göstermektedir. Gösterilen başarı yüzdeleri, farklı kişilerden alınan örnekler için ayrı ayrı hesaplanmıştır. Bu grafiklerde de gösterildiği gibi birinci algoritmada ortalama başarı %95,89 olarak hesaplanmıştır, ikinci ve beşinci veri kümesinde örtme olduğu için başarı düşmüştür. İkinci algoritmada ise ortalama başarı %98,41 olarak ölçülmüştür.

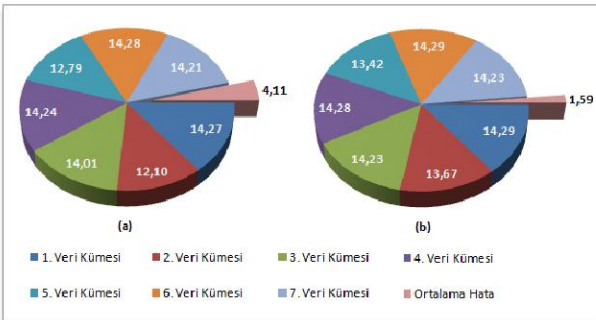


Şekil 4: Birinci izleme algoritmasının toplamdaki başarı yüzdeleri.



Şekil 5: İkinci izleme algoritmasının toplamdaki başarı yüzdeleri.

Şekil 6'da, farklı örnek setlerinin ortalama başarıya etkisi, tasarlanan iki farklı izleme algoritması için ayrı ayrı gösterilmiştir.



Şekil 6: (a) Birinci izleme algoritmasında, (b) ikinci izleme algoritmasında farklı örneklerin ortalama başarıya etkisi ve ortalama hata yüzdeleri.

8. Teşekkür

Yazarlar, veri kümesi için görüntü toplanmasında yardımcı olan Caner Aksoy, Mehmet Güney, Can Işın, Pınar Kavak, Nedret Özay ve Özlem Özcan'a teşekkür eder.

Kaynakça

- [1] Guil N, Zapata E, "Lower order circle and ellipse Hough transform," Pattern Recognition 30(10):1729-1744, 1997.
- [2] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Transactions of the ASME-Journal of Basic Engineering*, vol. 82, pp. 35-45, 1960.
- [3] I. K. Sethi and R. Jain, "Finding trajectories of feature points in a monocular image sequence," IEEE Trans. Pattern Analysis and Machine Intelligence , 9:56-73, 1987.
- [4] K. Rangarajan and M. Shah, "Establishing motion correspondence," CVGIP: Image Understanding , 54:56-73, 1991.
- [5] D. Chetverikov and J. Verestóy, "Motion Tracking of Dense Feature Point Sets," In Proc. 21st Workshop of the Austrian Pattern Recognition Group , pages 233-242. Oldenbourg Verlag, 1997.
- [6] L. R. Rabiner and B. H. Juang, "An introduction to hidden markov models," IEEE ASSP Magazine, pp. 4-16, January 1986.
- [7] M. Brand and A. Hertzmann, "Style machines," in SIGGRAPH 2000, July 2000, pp. 183-192.
- [8] Saber, E. and Tekalp, A., "Frontal-View Face Detection and Facial Feature Extraction using Color, Shape and Symmetry Based Cost Functions," *Pattern Recognition Letters* 19,8, 1998.
- [9] Vezhnevets, V., Sazonov, V. and Andreeva, A., "A Survey on Pixel-Based Skin Color Detection Techniques," *Proc. Graphicon-2003*, pp. 85-92, Russia, 2003.
- [10] Kovac, J., Peer, P. and Solina, F., "Human Skin Colour Clustering for Face Detection," *EUROCON '98 Computer as a Tool, Vol 2*, 2003.