カスケード型識別器を用いたパーティクルフィルタによる人物三次元追跡

佐藤洋一

情報理工学系研究科電子情報学専攻

1 はじめに

カメラからの入力画像を用いて人物を追跡する技術は，カメラの小型化や低価格化，防犯を目的とした監視カメラの普及などに伴い，セキュリティやマーケティングなどへの応用が期待されている．

カメラ画像を用いた対象の追跡は，これまでに多くの手法が提案されているが，なかでも近年，パーティクルフィルタ [2] の有効性が数多く報告されている ([1, 3, 4, 5, 6, 7, 8, 10, 11] など)．パーティクルフィルタは，状態量と尤度を持つ多数の仮説群により離散的な確率密度として追跡対象を表現し，それを状態遷移モデルを用いて伝播させることで，動きの変動や観測のノイズに対して頑健な追跡を実現する手法である．この手法は，観測値が非ガウス型になる状況においても頑健な追跡を実現することができるとして注目されている．

パーティクルフィルタによる人物追跡では，仮説の人物らしさをカメラ画像を用いて評価するが，これには，楕円と仮定した頭部のカラーヒストグラムや輪郭の輝度変化の類似性などが用いられることが多い ([1, 2, 3, 5, 7, 8, 10])．しかし，このような比較的単純な評価手法は，照明変動や複雑な背景下における人物の追跡では必ずしも十分ではなく，性能の向上には頑健かつ高精度な評価手法が求められる．

一方で，静止画像などから人物の顔を検出する手法が数多く提案されている．なかでも，ブースティング学習による識別器を用いた顔検出手法が良く知られており，特に，Haar-like 特徴を用いた AdaBoost ベース識別器による顔検出手法は，照明変動や，低解像度での観察にも比較的強く，実行速度の速さと検出精度から，広く利用されるようになってきた．AdaBoost ベース識別器は多数の弱識別器を線形結合することで高精度な識別器を構成するが，Viola と Jones はこれをカスケード型とし，学習，検出時に用いる Haar-like 特徴を高速に計算する手法 [9] を提案している．

カスケード型 AdaBoost ベース識別器は，学習に多くの時間を要するものの，識別時には非検出対象はカスケードの初期に棄却されるため，単一の識別対象に対しては高速な処理が可能である．しかし，対象を追跡するために，識別対象の大きさをさまざまに変化させながら，画面全体を逐次探索することは効率的ではない．そこで，パーティクルフィルタの仮説の評価にカスケード型 AdaBoost ベース識別器を応用し，探索範囲を限定することは有効であると考えられる．

このような知見に基づいて，本論文では，パーティクルフィルタの枠組みにおいて，仮説の評価に Haar-like 特徴を用いたカスケード型 AdaBoost ベース識別器を応用した人物追跡手法を提案する．追跡には視野を共有した複数のカメラを用い，三次元位置と向きを状態量として，一人の人物頭部を三次元空間内で追跡する．このとき，人物頭部は実際の向きやカメラとの位置関係により，その見えが変化するため，各方向の頭部の向き毎に学習を行った識別器を複数準備し，視野を共有した複数のカメラによりさまざまな方向で観察される人物頭部に対し，識別器を選択的に用いて評価する．

本論文に関連した手法としては，パーティクル

フィルタとカスケード型 AdaBoost ベース識別器を併用した Okuma らの手法 [6] が知られている．しかし，Okuma らの手法では，カスケード型 AdaBoost ベース識別器を主に画像からの対象検出に利用しており，パーティクルフィルタの仮説の評価に積極的に応用したものではない．

複数のカメラを用いて三次元空間内で追跡を行うものでは，複数のカメラによる仮説の評価を統合して人物を追跡する手法 [3, 10] や，さらに環境モデルを併用して頑健な追跡を行う手法 [8] が提案されているが，各カメラにおける仮説の評価には，比較的単純な手法が用いられるに留まっている．また，Nickel らの手法 [4] では，評価の一部にカスケード型 AdaBoost ベース識別器を利用しているが，対象の向きとカメラとの関係を考慮して，識別器を選択的に用いるものではない．これに対して本論文では，識別器を複数準備し，人物の実際の向きとカメラとの位置関係を考慮して，識別器を選択的に用いて評価する．

2 パーティクルフィルタ

時刻 t における対象の状態量を $\mathbf{x}_t$，画像による観測を $\mathbf{z}_t$ とし，時刻 t までに得られる観測を $\mathbf{Z}_t = \{\mathbf{z}_1, \ldots, \mathbf{z}_t\}$ とする．このとき，時刻 t における対象の事前確率 $P(\mathbf{x}_t \mid \mathbf{Z}_{t-1})$ は，マルコフ過程を仮定することにより，時刻 $t-1$ における事後確率 $P(\mathbf{x}_{t-1} \mid \mathbf{Z}_{t-1})$ と時刻 $t-1$ から t への状態遷移確率 $P(\mathbf{x}_t \mid \mathbf{x}_{t-1})$ を用いて以下のように表すことができる．

$$P(\mathbf{x}_t \mid \mathbf{Z}_{t-1}) = \int P(\mathbf{x}_t \mid \mathbf{x}_{t-1}) P(\mathbf{x}_{t-1} \mid \mathbf{Z}_{t-1}) d\mathbf{x}_{t-1} \quad (1)$$

ここで，$P(\mathbf{z}_t \mid \mathbf{Z}_{t-1})$ を一定とすると，時刻 t における事後確率 $P(\mathbf{x}_t \mid \mathbf{Z}_t)$ は，ベイズの法則より，時刻 t における尤度 $P(\mathbf{z}_t \mid \mathbf{x}_t)$ と事前確率 $P(\mathbf{x}_t \mid \mathbf{Z}_{t-1})$ により次式のように表すことができる．

$$P(\mathbf{x}_t \mid \mathbf{Z}_t) \propto P(\mathbf{z}_t \mid \mathbf{x}_t) P(\mathbf{x}_t \mid \mathbf{Z}_{t-1}) \quad (2)$$

対象の追跡は，この事後確率 $P(\mathbf{x}_t \mid \mathbf{Z}_t)$ の期待値を逐次求めることで実現される．

パーティクルフィルタでは，時刻 t における事後確率 $P(\mathbf{x}_t \mid \mathbf{Z}_t)$ を，状態量 $\mathbf{x}_t$ の仮説群 $\{\mathbf{s}_t^{(1)}, \ldots, \mathbf{s}_t^{(N)}\}$ と各仮説に対応する重み $\{\pi_t^{(1)}, \ldots, \pi_t^{(N)}\}$ により離散的に近似し，次のプロセスを経て，逐次的に更新する．

1. 仮説の選択

   時刻 $t-1$ における事後確率 $P(\mathbf{x}_{t-1} \mid \mathbf{Z}_{t-1})$ を離散的に近似した N 個の仮説 $\{\mathbf{s}_{t-1}^{(1)}, \ldots, \mathbf{s}_{t-1}^{(N)}\}$ の重み $\{\pi_{t-1}^{(1)}, \ldots, \pi_{t-1}^{(N)}\}$ の比に従い，仮説群 $\{\mathbf{s}_{t-1}'^{(1)}, \ldots, \mathbf{s}_{t-1}'^{(N)}\}$ を選択する．

2. 状態遷移確率に基づく伝播

   選択された仮説群 $\{\mathbf{s}_{t-1}'^{(1)}, \ldots, \mathbf{s}_{t-1}'^{(N)}\}$ を，状態遷移確率 $P(\mathbf{x}_t \mid \mathbf{x}_{t-1} = \mathbf{s}_{t-1}'^{(n)})$ に従い伝播し，$P(\mathbf{x}_t \mid \mathbf{Z}_{t-1})$ に相当する時刻 t における N 個の仮説群 $\{\mathbf{s}_t^{(1)}, \ldots, \mathbf{s}_t^{(N)}\}$ を生成する．

3. 画像による重み $\pi_t^{(n)}$ の推定

   仮説群 $\{\mathbf{s}_t^{(1)}, \ldots, \mathbf{s}_t^{(N)}\}$ の重み $\pi_t^{(n)} \approx P(\mathbf{z}_t \mid \mathbf{x}_t = \mathbf{s}_t^{(n)})$ を尤度の評価を行うことで画像から推定する．

3 カスケード型 AdaBoost ベース識別器

Viola と Jones により提案されたカスケード型 AdaBoost ベース顔検出器 [9] は，検出時間の短縮のため，複数の識別器を組み合わせたカスケード構造をなしている．入力画像に対し，各段で顔，非顔の判定を行い，顔と判定された画像だけが次の段へ進む．最後の段まで通過したものが最終的に顔と判定される．

カスケードの各段において，学習用顔画像を通過させる割合 (学習用顔画像通過率) を $Dr(0 < Dr < 1)$，学習用非顔画像を通過させる割合 (学習用非顔画像通過率) を $Fp(0 < Fp < 1)$ とすると，カスケード n 段通過後は，学習用顔画像は Dr^n，学習用非顔画像は Fp^n だけ通過していることになり，例えば $Dr = 0.999$，$Fp = 0.5$ とした

場合，$n = 40$ のカスケード型識別器は学習用顔画像通過率は $0.999^{40} \approx 0.96$，学習用非顔画像通過率は $0.5^{40} \approx 9.1 \times 10^{-13}$ となり，学習用顔画像をほとんど通過させ，学習用非顔画像をほとんど通過させない顔検出器となる．

カスケードの各段を構成する識別器 $\mathbf{H}_i(x)$ は多数の弱識別器 $h_t(x)$ の線形結合により，以下のように表される．

$$\mathbf{H}_i(x) = sgn\left(\sum_{t=1}^{T} \alpha_t h_t(x)\right) \tag{3}$$

ここで，T は用いられる弱識別器の数であり，α_t は学習時に決まる弱識別器のエラー ϵ_t を用いて $\alpha_t = \log \frac{1-\epsilon_t}{\epsilon_t}$ と表される．

特徴矩形の位置と大きさを画像内でどのようにとるかによって膨大な種類の特徴が存在するが，これらの中から顔をよく識別する特徴が学習時に AdaBoost アルゴリズムにより選択され，各段の識別器が準備される．

4 提案手法

視野を共有した複数のカメラを用いて，三次元位置と向きを状態量として，一人の人物頭部を，パーティクルフィルタにより追跡する手法を提案する．本論文が新たに提案する手法は以下である．1) カスケード型 AdaBoost ベース識別器をパーティクルフィルタの仮説の評価へ積極的に応用し，2) 各方向の頭部の向き毎に学習を行った識別器を，仮説とカメラの関係に基づいて選択的に用いることで，複数のカメラでの仮説の評価を可能とする．以下に，提案手法について詳細を述べる．

4.1 人物頭部モデルと仮説のカメラ画像への射影

室内空間に三次元世界座標系 XYZ をとる．座標系は床面を XY 平面と一致させ，高さ方向を Z 軸とする．人物頭部はモデルとして楕円体を仮定する．このとき，人物頭部は一定の大きさの剛体とし，位置を楕円体の中心座標 (x, y, z) で表現する．また，人物は頭部を傾けて室内を移動することは少ないと仮定すると，人物頭部の向きは，X 軸を基準とした Z 軸回りの回転 θ のみで表せる．

次に，人物頭部の時刻 t における n 番目の仮説 $\mathbf{s}_t^{(n)} = \left[x_t^{(n)}, y_t^{(n)}, z_t^{(n)}, \theta_t^{(n)}\right]^\top$ は，校正済みの i 番目のカメラ画像に次のように射影することができる．

$$\mathbf{p}_{i,t}^{(n)} = F_i\left(\mathbf{s}_t^{(n)}\right) \tag{4}$$

ここで $\mathbf{p}_{i,t}^{(n)}$ は，仮説 $\mathbf{s}_t^{(n)}$ の位置を i 番目のカメラ画像へ射影したものであり，カメラ画像座標 $\left[u_{i,t}^{(n)}, v_{i,t}^{(n)}\right]^\top$ を要素に持つ．

また，各カメラによって観察される相対的な人物頭部の向きは以下のように表される．

$$\theta_{i,t}^{(n)} = \theta_t^{(n)} - tan^{-1}\left(\frac{\left[\mathbf{J}\mathbf{c}_i - \mathbf{K}\mathbf{s}_t^{(n)}\right]^y}{\left[\mathbf{J}\mathbf{c}_i - \mathbf{K}\mathbf{s}_t^{(n)}\right]^x}\right) \tag{5}$$

ここで $\theta_{i,t}^{(n)}$ は i 番目のカメラによって観察される相対的な人物頭部の向きである．$\mathbf{J}$ はカメラ位置 $\mathbf{c}_i$ から XY 位置成分を取り出すための行列であり，$\mathbf{K}$ は仮説 $\mathbf{s}_t^{(n)}$ から XY 位置成分を取り出すための行列である．$[\,]^x$ は計算結果から X 軸に対応する要素を取り出すことを表している．

各カメラで観察される人物頭部の幅 (l_i) は，人物頭部の楕円体モデルを射影したものを用いる．

4.2 カスケード型 AdaBoost ベース識別器による人物頭部らしさの評価

カスケード型 AdaBoost ベース識別器を画像に適用する場合，通常，画像中の対象周辺に発生する多くの候補領域をマージして最終的に一つの対象として検出する．しかし，本手法では，マージに相当する処理はパーティクルフィルタの枠組みにより提供されるため，カスケード型 AdaBoost ベース識別器による評価は，各仮説に対応した人物頭部候補領域画像 $g_{i,t}^{(n)}$ について独立して考えれば良い．

カスケードの各段の識別器は，階層が進むにしたがって，より多くの弱識別器 $h_t(x)$ を用いて判定を行う．そのため，より多くの識別器を通過した人物頭部候補領域画像 $g_{i,t}^{(n)}$ は，より多くの人物頭部の特徴を保持している．このような知見に基づいて，本手法では人物頭部候補領域画像 $g_{i,t}^{(n)}$ をカスケード型 AdaBoost ベース識別器に入力した際に通過した識別器の数 (カスケード段数) を人物頭部らしさの評価とする．これは，仮説が実際の人物頭部の状態と大きく離れて生成された場合，対応する人物頭部候補領域画像 $g_{i,t}^{(n)}$ はカスケードの初期に棄却されるため，計算コストの点からも都合が良い．

このような考えに基づいて，時刻 t における n 番目の仮説 $\mathbf{s}_t^{(n)}$ を i 番目のカメラに射影した際の重み $\pi_{i,t}^{(n)}$ は，以下の手順により得る．ただし事前に，カスケード型 AdaBoost ベース識別器を，正面，90° 左向き，90° 右向きなどの人物頭部の向き毎に，人物頭部と非人物頭部で通過する識別器の数 (カスケード段数) に十分な差がつくように学習しておく．

1. 各時刻 t において，生成された n 番目の仮説 $\mathbf{s}_t^{(n)}$ を i 番目のカメラ画像に射影し，カメラ画像座標 $\mathbf{p}_{i,t}^{(n)}$，相対的な人物頭部の向き $\theta_{i,t}^{(n)}$，カメラ画像上での人物頭部の幅 $l_{i,t}^{(n)}$ を得る．

2. 仮説 $\mathbf{s}_t^{(n)}$ を射影したカメラ画像座標 $\mathbf{p}_{i,t}^{(n)}$ を中心に，カメラ画像上での人物頭部の幅 $l_{i,t}^{(n)}$ を一辺とする領域を切り出す．ただし，仮説を射影した際にカメラの視野外となる場合，評価は行わず，重みを一定の小さな値とする．

3. カスケード型 AdaBoost ベース識別器は，識別対象画像サイズが固定 (例えば 24×24 ピクセルなど) であるため，(2) で切り出した画像のサイズを変更し，識別器に入力可能な人物頭部候補領域画像 $g_{i,t}^{(n)}$ を生成する．

4. 仮説の射影によって得られた相対的な人物頭部の向き $\theta_{i,t}^{(n)}$ に基づいて，カスケード型 AdaBoost ベース識別器を選択する．例えば，正面，90° 右向き，90° 左向きの 3 方向の識別器を用いた場合，相対的な人物頭部の向き $\theta_{i,t}^{(n)}$ が −45°〜45° の場合は人物頭部正面の識別器が選択され，45°〜135° の場合は 90° 左向きの識別器が選択され，−45°〜−135° の場合は 90° 右向きの識別器が選択される．

5. 人物頭部候補領域画像 $g_{i,t}^{(n)}$ を選択されたカスケード型 AdaBoost ベース識別器に入力し，人物頭部候補領域画像 $g_{i,t}^{(n)}$ が通過した識別器の数 (カスケード段数) を取得する．ここで得たカスケード段数を対応する仮説の重み $\pi_{i,t}^{(n)}$ とする．例えば，全カスケード段数が 40 段で，すべての識別器を通過した場合，重みは 40 となり，5 段目で棄却された場合，重みは 5 となる．

4.3 尤度の統合

仮説 $\mathbf{s}_t^{(n)}$ を各カメラ画像に射影し，カスケード型 AdaBoost ベース識別器に入力して得た重み $\pi_{i,t}^{(n)}$ を統合する．重み $\pi_{i,t}^{(n)}$ の統合は，次式のように各カメラによる人物頭部らしさの評価に基づく重みの積とし，期待値をとることで各時刻の人物頭部の状態量を推定する．

$$\pi_t^{(n)} = \prod_i \pi_{i,t}^{(n)} \tag{6}$$

5 実験

これまでの提案手法に基づき，人物頭部追跡の実験を行った結果について述べる．実験は室内天井に IEEE1394 カラーカメラ (Point Grey Research 社製 Flea) を設置して行った．各カメラの映像は 640×480 ピクセル，毎秒 30 フレームの画像列で得ることとし，校正済みの 2 台のカメラの映像を 1 台の汎用 PC(CPU Intel Petium4 3.2GHz, Memory 1GByte, OS WindowsXP) で処理した．実験で利用したカメラから得られた画像の例を図 1 に示す．

カスケード型 AdaBoost ベース識別器は 3 種とし，人物頭部の正面，90° 右向き，90° 左向きをそれぞれ検出するよう学習を行った．カスケード

(a) Camera 1　　(b) Camera 2

図 1: カメラ画像の例

段数は 40 段とし，識別対象画像サイズは 24×24 ピクセルとした．また，パーティクルフィルタの仮説数は 200 とし，状態遷移モデルとして等速直線運動を仮定した．

頭部の向きを変えながら観測領域内を移動する人物を追跡する実験を行なった。

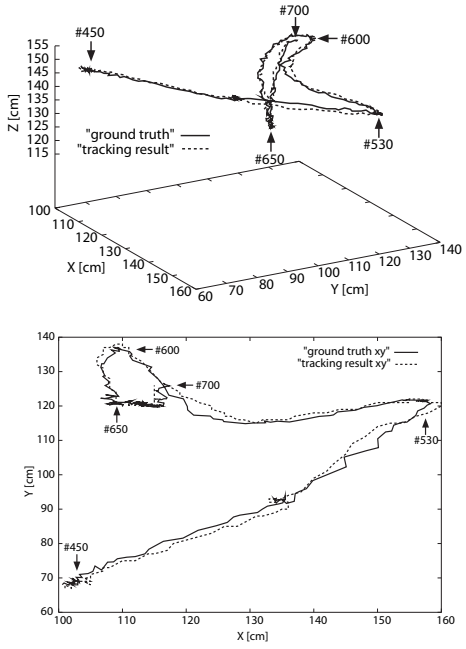


図 2: 人物頭部の追跡結果の軌跡

提案手法による追跡の精度を定量的に調べるために，画像中の人物頭部位置を手作業で求め, 2 枚の画像から逆投影して求めた三次元座標を真の位置と見なし，推定結果と比較した．図 2 に，推定結果と対応する人物頭部の真の位置の三次元空間上，及び XY 平面上での軌跡を示す．XY 平面上の平均誤差，Z 軸方向の平均誤差は共に 2cm 以内であり，高精度な追跡ができているといえる．

なお，実験に用いた PC では，200 個の仮説を 15ms 程度で処理できた．2 台のカメラに仮説を射影した際の処理は合計 30ms 程度で完了し，リアルタイムでの追跡が可能であった．

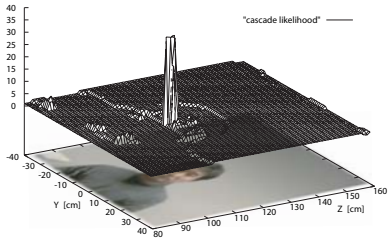
パーティクルフィルタを用いた追跡では，各フレームにおいて，仮説の尤度を高精度に推定できることが性能向上に大きく寄与する．本手法の枠組みにおいては，人物頭部の仮説を画像平面上に射影した際に，人物頭部周辺で鋭いピークを持つ関数が理想的である．そこで，実際の人物頭部の位置周辺で，カメラから一定距離の平面を一辺 1cm のグリッドで分割し，それぞれの三次元位置での人物頭部らしさの評価を提案手法に基づいて算出した．その結果を図 3(a) に示す．

図 3 より，カスケード段数に基づく評価が頭部周辺で高く，頭部以外の場所では低くなっていることが分かる．また，輪郭の輝度変化の類似性に基づく評価に比べ，頭部周辺での評価が鋭いピークをもっており，パーティクルフィルタでの利用に適しているといえる．

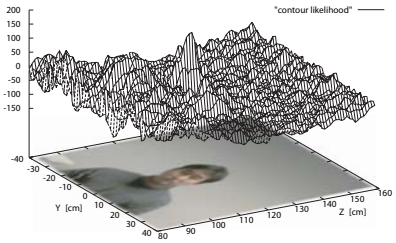
6　おわりに

本論文では，パーティクルフィルタにおける仮説の評価に，カスケード型 AdaBoost ベース識別器を応用し，実時間で人物頭部の追跡を行う手法を提案した．複数の識別器を仮説に基づいて選択的に用いることで，人物が頭部の向きを変えながら移動しても，高精度に人物頭部を追跡できることを示した．

また，本手法によって，人物頭部の向きを測定することが可能であるが，正面，90° 右向き，90° 左向きの 3 方向の識別器を用いた実験では，推定される顔の向きに 45° 程度のばらつきがあった．これに対しては，識別器の方向を追加することで，より精度良く推定することができると予想される．なお，本手法では，パーティクルフィルタの仮説により選択的に識別器が用いられるため，計算コ



(a) カスケード段数に基づく評価



(b) 輪郭の輝度変化の類似度に基づく評価

図 3: 評価の分布

ストを増加させずに識別器の種類を追加することが可能である.

今後は，より細かな方向の識別器を追加した場合の追跡精度の検討や，複数人物追跡への拡張，頭部の初期検出などについて検討する予定である.

参考文献

[1] S. Birchfield, “Elliptical Head Tracking Using Intensity Gradients and Color Histograms,” Proc. the IEEE International Conference on Computer Vision and Pattern Recognition, pp.232-237, 1998.

[2] M. Isard and A. Blake, “Condensation - Conditional Density Propagation for Visual Tracking,” International Journal of Computer Vision, vol.29, no.1, pp.5-28, 1998.

[3] 松本郁佑, 加藤丈和, 和田俊和, “Network Augmented Multisensor Assosiation-CONDENSATION: CONDENSATION の自然な拡張による 3 次元空間内での人物頭部の実時間追跡,” 情報処理学会研究報告, 2005-CVIM-150-21, pp.161-168, 2005.

[4] K. Nickel, T. Gehrig, R. Stiefelhagen and J. McDonough, “A Joint Particle Filter for Audio-visual Speaker Tracking,” Proc. the 7th international conference on Multimodal interfaces, pp.61-68, 2005.

[5] K. Nummiaro, E. Koller-Meier and L. Van Gool, “An Adaptive Color-Based Particle Filter,” Image and Vision Computing, vol.21, no.1, pp.99-110, 2003.

[6] K. Okuma, A. Taleghani, N. Freitas, J. Little and D. Lowe, “A Boosted Particle Filter: Multitarget Detection and Tracking,” European Conference on Computer Vision, vol.3021 of LNCS, pp.28-39, 2004.

[7] 杉本晃宏, 谷内清剛, 松山隆司, “確信度付き仮説群の相互作用に基づく複数対象追跡,” 情報処理学会論文誌, vol.43 no.SIG CVIM 4, pp.69-84, 2002.

[8] 鈴木達也, 岩崎慎介, 小林貴訓, 佐藤洋一, 杉本晃宏, “環境モデルの導入による人物追跡の安定化,” 電子情報通信学会論文誌 DII, vol.J88-DII no.8, pp.1592-1600, 2005.

[9] P. Viola and M. Jones, “Rapid Object Detection Using a Boosted Cascade of Simple Features,” Proc. the IEEE International Conference on Computer Vision and Pattern Recognition, vol.1, pp.511-518, 2001.

[10] Y. Wang, J. Wu and A. Kassim, “Particle Filter for Visual Tracking Using Multiple Cameras,” Proc. IAPR Conference on Machine VIsion Applications, pp.298-301, 2005.

[11] C. Yang, R. Duraiswami and L. Davis, “Fast Multiple Object Tracking via a Hierarchical Particle Filter,” Proc. the IEEE International Conference on Computer Vision and Pattern Recognition, vol.1, pp.212-219, 2005.