Analyse topographique de cartes de similarité dans l'évaluation de performances pour le suivi d'objets

Mounia Mikram Rémi Mégret Yannick Berthoumieu

Laboratoire IMS, UMR CNRS 5218, Université de Bordeaux

{mounia.mikram, remi.megret, yannick.berthoumieu}@ims-bordeaux.fr

Résumé

Cet article présente une nouvelle approche quantitative visant à caractériser les performances d'un modèle d'apparence dans le contexte du suivi d'objets. Les performances sont calculées à partir de la topographie de cartes de similarité obtenues en comparant une référence et l'ensemble des régions obtenues en faisant varier le paramètre de position dans l'image courante. La distance spatiale entre la position vraie de l'objet et plusieurs positions optimales au sens de cette carte permet de caractériser les performances du modèle d'apparence de façon indépendante du type de similarité utilisée, ce qui permet la comparaison objective des modèles. Des résultats obtenus sur des vidéos réelles illustreront l'intérêt de l'approche. Cette approche se veut complémentaire de l'approche classique consistant à évaluer un système de suivi entier sur la base des estimations qu'il produit, en se focalisant sur les performances intrinsèques d'un modèle d'apparence.

Mots clefs

Evaluation de performances, modèle d'apparence, suivi d'objets, descripteur visuel, carte de similarité.

1 Introduction

1.1 Contexte

Différentes méthodes pour la mesure des performances de systèmes de suivi ont déjà été proposées [1,2,3,4]. Chacune de ces méthodes évalue les performances grâce à un certain nombre de mesures sur la qualité de la localisation estimée par le système. Ces mesures se fondent sur un corpus vidéo, par exemple [5] auquel est associée une vérité terrain qui capture l'interprétation vraie de la scène en termes d'objets à suivre.

Une telle évaluation prend en compte uniquement la réponse fournie par le système, ce qui correspond à une approche de type "boîte noire" (Fig. 1). Ce type d'évaluation, même si elle offre une quantification utile des performances, cantonne la mesure à un niveau global et ne permet pas de caractériser les performances intrinsèques des différents éléments composant le système.

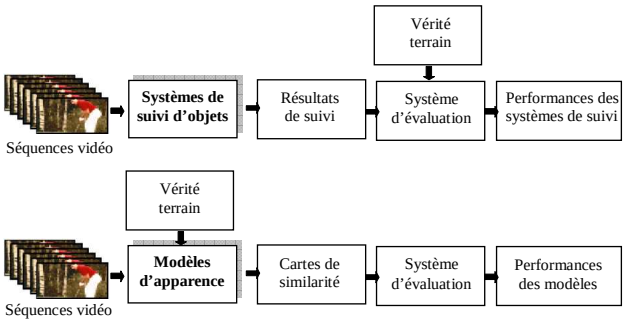


Fig 1. Paradigme standard d'évaluation des systèmes de suivi d'objet (en haut), et paradigme proposé pour l'évaluation des modèles d'apparence (en bas).

À notre connaissance, très peu de travaux ont proposé une étude quantitative des performances comparées de modèles d'apparence séparément de l'algorithme de recherche. La plupart des travaux se limitent en effet à une illustration des performances en termes de suivi ou à l'étude qualitative de la fonction de similarité, dans le contexte de la proposition de nouveaux algorithmes.

1.2 Travaux antérieurs

La performance d'une mesure de similarité dans la mise en correspondance des images en niveaux de gris a été étudiée dans les travaux de Sohal et al. [6]. Ces auteurs ont analysé en particulier l'utilisation d'histogrammes de niveaux de gris comparés par le coefficient de Bhattacharyya et la divergence de Kullback-Leibler. Ils ont montré que ces deux mesures donnent des estimations biaisées sur la localisation des objets dans une séquence vidéo avec des images en niveaux de gris.

Dans leurs travaux, la carte de la similarité de l'erreur quadratique moyenne des différences pixel à pixel (EQM) est considérée comme référence et utilisée pour analyser les cartes de similarité du coefficient Bhattacharyya et Kullback Leibler.

Cette analyse s'appuie sur le maximum de la carte de similarité qui fournit la position de l'objet où la meilleure correspondance se produit. Un pic étroit et correctement positionné indique une bonne mise en correspondance avec la cible, alors qu'un pic large fait apparaître une ambigüité sur l'estimation de la position de l'objet cible, qui peut conduire à une localisation imprécise.

L'étude de la carte de similarité ainsi proposée s'intéresse à la précision finale obtenue en considérant uniquement l'optimum de la carte de similarité. Ceci fournit une information utile dans le cas d'une recherche exhaustive, mais un modèle d'apparence est destiné à être inclus dans un système plus vaste, qui peut comprendre des algorithmes de recherche variés. Nous proposons dans la suite de compléter le paradigme basé sur l'étude de la carte de similarité (Fig. 1), à l'aide de plusieurs métriques qui capturent des aspects variés de la performance d'un modèle d'apparence. L'objectif de cet article est d'offrir des outils pour une telle analyse.

2 Principe général

2.1 Carte de dissimilarité

Dans un système de suivi fondé sur une représentation par boîte englobante, on peut définir l'état $\boldsymbol{\theta}$ de l'objet comme étant le vecteur composé des coordonnées de cette boîte englobante. L'objet à suivre dans une image de référence I_{ref} est ainsi représenté par une boîte $\mathbf{b}^*_{n,t}$. A l'intérieur de cette boîte est calculé un descripteur caractérisant l'apparence de l'objet. Pour accomplir la tâche de suivi le long de la séquence, le système de suivi est généralement muni d'une architecture qui comprend trois parties :

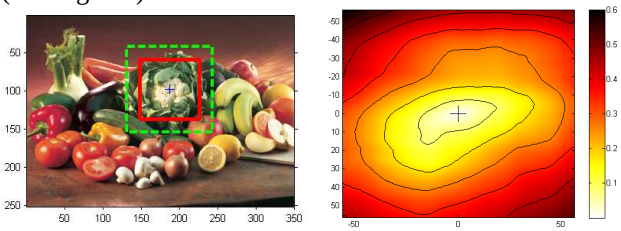
- le modèle d'apparence, qui décrit ce à quoi un objet doit ressembler dans une image. Ce modèle peut être représenté par un couple (descripteur, similarité) ;
- l'algorithme d'optimisation, qui tente d'estimer la position de l'objet en optimisant la correspondance entre l'apparence courante et le modèle d'apparence de référence ;
- les contraintes spatio-temporelles sur le mouvement de l'objet, permettant de restreindre l'espace de recherche ou de favoriser les trajectoires les plus vraisemblables (notamment sur la base d'une régularité du mouvement).

L'étude présente se focalise sur la caractérisation du modèle d'apparence. Pour comparer les performances de plusieurs modèles d'apparence, nous proposons d'analyser de façon plus quantitative leurs cartes de similarités.

Une carte de similarité est définie pour un modèle d'apparence, une image courante, et une référence. En premier lieu, une fenêtre rectangulaire englobant l'objet d'intérêt est positionnée sur une image référence. Cette fenêtre est associée au modèle cible, alors qu'une autre fenêtre de recherche dans un voisinage de l'objet d'intérêt sera déterminée dans une image cible.

Plus formellement, soit $\{\mathbf{x}_i^j = [x_i^j, y_i^j]^T\}_{i=1,..,n;\, j=1}$ la région de référence (rectangle rouge, fig. 2.a) entourée d'un ensemble de F régions candidates $\{\mathbf{x}_i^j = [x_i^j, y_i^j]^T\}_{i=1,..,n;\, j=1,..,F}$ situées dans une fenêtre de recherche (rectangle vert, fig. 2.a).

Les descripteurs utilisés sont calculés pour chaque région. Les régions candidates sont comparées de manière exhaustive avec la région référence en utilisant une mesure de similarité générant ainsi une carte de similarité (voir fig. 2.b)



(a) (b)

Fig 2. Processus de génération d'une carte de disimilarité. (a) : la région référence est marquée en rouge et les candidats sont calculés dans la région pointillée en vert, (ici image référence = image cible). (b) : la carte de disimilarité est calculée en utilisant la distance de Bhattacharyya entre les histogrammes de la région de référence et des régions candidates.

2.2 Analyse topographique

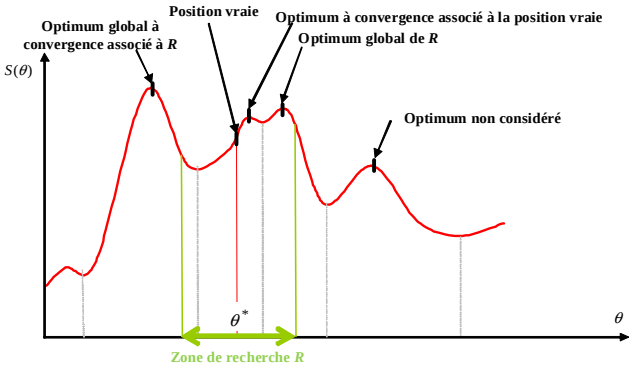


Fig 3. Définition des éléments topographiques calculés sur la carte de similarité.

L'analyse topographique nous permet d'établir des mesures pour étudier la performance d'un modèle d'apparence. Les deux entrées de l'étude sont :

– la position vraie θ^*,
– la carte de similarité $S(\theta)$ (on pourra sans perte de généralité considérer à sa place une carte de dissimilarité $D(\theta)$, en inversant l'axe des valeurs).

L'ensemble des mesures proposées ci-dessous ont pour objectif de caractériser la précision et l'ambiguïté de l'estimation de l'optimum de la carte de similarité au voisinage de la position vraie θ^*. Pour ce faire, nous proposons de considérer une zone de recherche R, centrée en θ^*, et de rayon spatial arbitraire r. Cette zone est considérée comme un ensemble de positions

envisageables pour l'initiation d'une recherche d'optimum.

Le principe que nous allons développer pour la caractérisation de la qualité d'un modèle d'apparence est d'étudier la forme de la carte de similarité au voisinage de cette zone de recherche, afin d'en déduire des caractéristiques mettant en évidence la capacité de détecter de façon précise et non ambigüe la position vraie uniquement sur la base de la similarité. On déduit ainsi de la carte de similarité un certain nombre d'éléments (voir fig. 3).

L'*optimum global sur la région R*, $\hat{\theta}^R = \arg\max_{\theta \in R} S(\theta)$ correspond à la meilleure similarité parmi toutes les positions testées, lors d'une recherche exhaustive dans un voisinage de la position vraie.

L'optimum local $\theta(x)$ associé à une position d'origine x arbitraire est défini comme l'optimum auquel est associé le bassin d'attraction contenant x, c'est-à-dire la position estimée par approche ascendante à partir de x. Nous l'appellerons optimum à convergence associé à x.

Lorsque l'on considère l'ensemble des positions de la zone de recherche, celles-ci sont associées à un ensemble $\{ \theta_{R,1} , \theta_{R,2} , \ldots \theta_{R,n}\}$ de $n \geq 1$ optima locaux, que nous appellerons ensemble des optima à convergence, qui correspondent aux optima dont les bassins versants ont une intersection non nulle avec la zone de recherche.
Parmi l'ensemble des optima à convergence associés à la région R, nous nous intéresserons plus particulièrement à deux d'entre eux :

- L'*optimum à convergence au pire cas* correspond à l'optimum de l'ensemble qui est situé le plus loin en distance spatiale de la position vraie. Nous le noterons $\theta_{R,P}$. Il correspond à la plus mauvaise estimation possible de la position, lorsque l'on initialise un algorithme de recherche de type ascendante sur l'une des positions de R.
- L'*optimum à convergence à meilleure similarité* correspond à l'optimum de l'ensemble qui a la meilleure similarité. Nous le noterons $\theta_{R,S}$. Il correspond à l'estimation qui serait jugée la meilleure par une recherche locale d'optimum, en testant toute les initialisations possibles sur la région R.

Les éléments précédents fournissent pour un rayon r donné trois positions, qui permettent de caractériser la précision de l'estimation dans trois types d'approches de recherche d'optimum différentes. Le choix de r a cependant une influence sur la précision mesurée.

Si l'on considère une région de rayon r faible, l'étude de la carte dans un tel voisinage peut renseigner principalement sur la capacité de la carte à posséder un optimum local proche de la position vraie.

Si l'on considère une région de rayon r plus important, cette étude peut se compléter de la détection d'ambiguïtés se caractérisant par la présence d'autres optima locaux éloignés de la position vraie, mais présentant néanmoins soit une bonne similarité, soit un chemin de remontée de gradient depuis une position au voisinage de la position vraie. En effet, si le point d'initialisation pour l'optimisation est situé dans un bassin d'attraction associé à un optimum éloigné, alors la cible sera perdue puisque l'algorithme convergera vers un optimum local incorrect qui faussera l'estimation de la position de l'objet suivi.

Fixer un unique rayon r arbitraire est difficile, la précision et la robustesse de l'estimation étant deux notions liées. En effet, la présence d'un optimum proche de la position vraie ne garantit pas que cet optimum soit choisi par l'algorithme de recherche, notamment dans le cas où d'autres optima sont également présents dans un voisinage proche. Pour cette raison, nous avons choisi de représenter les performances comme une fonction de r, ce qui permet de capturer à la fois la capacité à localiser avec précision et robustesse pour r faible, ainsi que la capacité à ne pas être ambigu pour r plus important. L'échelle d'analyse n'est ainsi pas fixée arbitrairement, mais s'adapte en fonction des données utilisées.

3 Résultats

Afin d'illustrer les méthodes d'évaluation de performance présentées, l'ensemble des résultats intermédiaires est présenté pour trois méthodes, sur un exemple caractéristique dans la section 3.1 puis des résultats quantitatifs sur un corpus d'évaluation seront présentés la section 3.2. Les images utilisées sont issues des vidéos publiquement disponibles issues du projet CAVIAR [5].

3.1 Etude d'un cas particulier

L'approche proposée est ici illustrée sur un couple d'images (la référence et l'image courante), montrées dans la figure fig. 4. Dans l'image de référence, la boîte englobante de l'objet de référence est indiquée en rouge. Dans l'image courante, cette boîte correspond à la région candidate pour une perturbation nulle, et est présentée en vert. Les cartes de distance entre l'objet de référence et les régions candidates sont présentées à la figure fig. 5.

Fig 4. Image de référence (à gauche) et cible (à droite). La vérité terrain de l'objet considéré est indiquée par un rectangle blanc. La carte de dissimilarité est calculée

pour toutes les régions dont le centre appartient à la région indiquée en pointillés jaunes.

Trois méthodes ont été testées ici : une méthode par histogrammes de couleurs sur la boîte englobante (CH pour *Color Histogram*), une méthode par histogrammes de couleurs pondérés spatialement à l'aide d'un noyau d'Epanechnikov adapté au rectangle englobant (WCH pour *Weighted Color Histogram*), et une méthode par imagette couleur 20×20 obtenue par rééchantillonnage du contenu de la boîte englobante (CT pour *Color Template*).

Les cartes de similarité obtenues sont montrées à la figure 5. En jaune clair, les lignes de partage des eaux de la carte permettent de discerner les bassins versants associés à chaque optimum local. Tout algorithme de type ascendant (tel que le suivi par Mean-Shift [7]), initialisé dans l'un de ces bassins versants converge vers l'optimum associé à ce bassin versant.

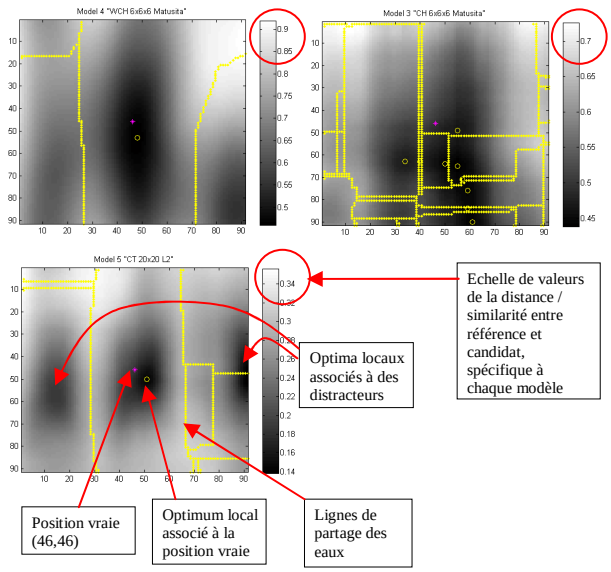


Fig 5. Exemple de cartes de (dis)similarité obtenues pour trois modèles d'apparence : CH (en haut), WCH (au milieu), CT (en bas).

Les cartes de similarité précédentes sont résumées sous la forme des trois types de courbes fonctions du rayon *r* de la zone de recherche, présentées dans les figures 6 et 7. Les courbes de la fig. 6 sont représentatives des performances attendues pour un algorithme balayant la zone de recherche de rayon r afin de trouver un paramètre associé à la meilleure similarité.

Les courbes de la fig. 7 sont représentatives des performances attendues lorsqu'une étape de remontée de gradient locale est utilisée. Elles illustrent ainsi la précision d'estimation pour *r* faible, qui indique la distance entre la position correcte et l'optimum qui lui est associé. D'autre part, la croissance de la courbe indique à partir de quel rayon un optimum plus éloigné risque d'être atteint si l'initialisation est imprécise. Il apparaît ici que les modèles WCH et CT sont relativement robustes jusqu'à une initialisation à 20 pixels de la position vraie, ce qui n'est pas le cas du modèle CH.

Les caractéristiques utilisées ne font pas intervenir de comparaisons entre similarités différentes, mais seulement des mesures sur des éléments calculés indépendamment sur chaque carte de similarité. Ceci permet de comparer deux modèles d'apparence directement au niveau des courbes de performance générées, malgré le fait que les composantes internes au modèle que sont les descripteurs ou la similarité utilisés soient de natures différentes.

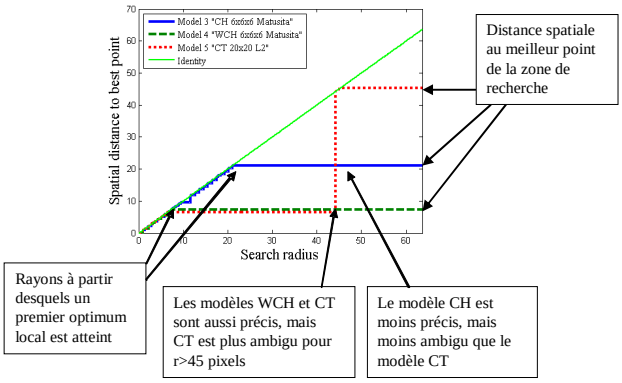


Fig 6. Courbes de distances à l'optimum global sur la région R

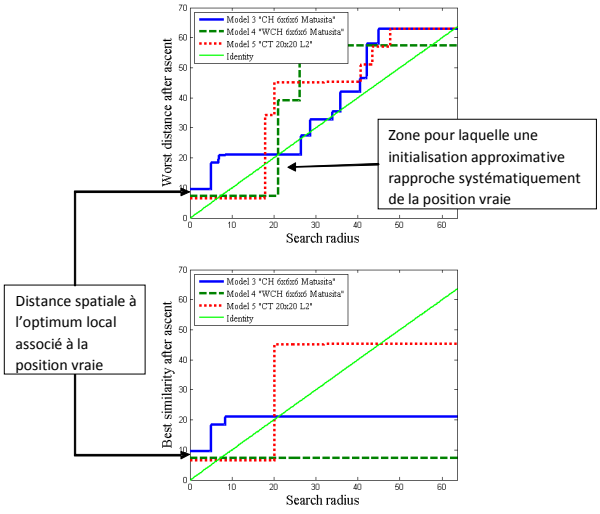


Fig 7. Courbes de distances à convergence : au pire cas (en haut) et à meilleure similarité (en bas)

3.2 Etude quantitative

Les expériences dans cette section ont été réalisées à partir de deux séquences vidéo issues du corpus CAVIAR [5]. Ces séquences sont de type vidéo surveillance prises avec une caméra fixe, montrant des êtres humains effectuant diverses actions. Les séquences viennent avec l'information de vérité terrain décrite par les boîtes

englobantes entourant les différents humains dans chaque scène.

Nous utilisons pour nos tests 2 séquences de la surveillance des centres commerciaux : 'ThreePastShop2cor' et 'OneShopOneWait2cor' (figure fig. 8). Nous allons étudier les performances des modèles d'apparence histogramme et template en utilisant 8 objets extraits de ces deux séquences. Les cartes de similarité sont calculées dans une fenêtre de recherche de 45×45 pixels avec un pas spatial de 4 pixels. L'image de référence ainsi que l'image candidate sont issues d'un échantillonnage de 40 images avec un pas temporel de 5 images.

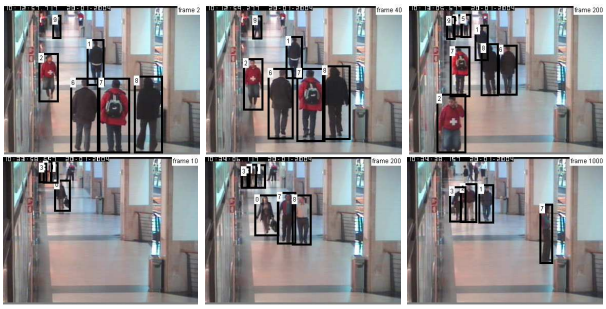


Fig 8. Quelques images de deux séquences extraites du projet CAVIAR [5] *pour illustrer l'approche proposée, associées à leur vérité terrain.*

Nos simulations montrent qu'il existe des objets où le bassin de convergence d'histogramme et plus grand que celui associé au template (fig. 10 à gauche) et d'autre cas où ils peuvent avoir des comportements plus proches (fig. 10 à droite). La figure 9 montre les courbes de distances à convergence moyennes pour l'histogramme et le template.

La précision est légèrement meilleure pour les approches template, ce qui se traduit pour les valeurs faibles de *r*. Ceci confirme les résultats de [6] indiquant une estimation biaisée de la part des modèles basés histogrammes de couleur, mais sans donner un rôle différent au modèle template qui est considéré comme un modèle au même niveau que les autres.

De plus, ces courbes révèlent qu'en moyenne les approches par histogramme ont un bassin versant plus grand que les approches par template, ce qui se traduit par une croissance plus lente de la distance à convergence au pire cas pour $r>3$. Ceci permet de quantifier la marge acceptable pour une initialisation conduisant pour chaque modèle à la convergence vers l'optimum correct.

4 Conclusion

Dans cet article, nous avons présenté une nouvelle approche pour l'évaluation des performances de modèles d'apparence pour le suivi d'objet. Les modèles sont supposés composés d'une étape d'extraction d'un descripteur sur la boîte englobante placée sur l'objet, et d'une étape de comparaison de ce descripteur avec une référence. La carte de similarité obtenue en balayant une zone de recherche caractérise la capacité du modèle à fournir une information à la fois précise et robuste de la position la plus vraisemblable. Les mesures basées sur l'analyse de la topographie de la carte permettent d'en extraire une évaluation quantitative, tout en autorisant la comparaison objective de modèles distincts.

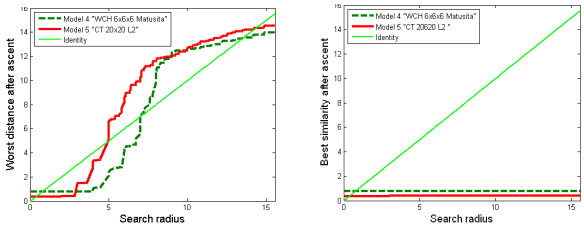


Fig 9. Courbes de distances à convergence pour plusieurs objets : au pire cas (à gauche) et à meilleure similarité (à droite).

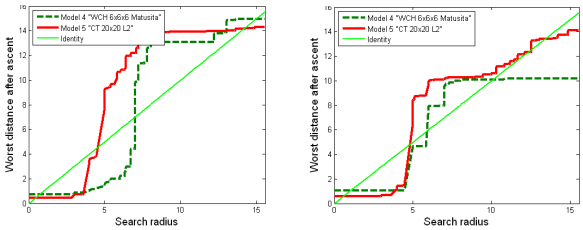


Fig 10. Courbes de distances à convergence au pire cas pour deux objets différents : (à gauche) l'histogramme à un bassin versant plus grand que le template (à droite).

Références

[1] J. Black, T. Elis et P. Rosin, "A novel method for video tracking performance evaluation,'' VS-PETS 2003, Nice, pp. 125–132.

[2] S.M Schneiders, T. Jager, H.S. Loos et W. Niem, "Performance Evaluation of a Real Time Video Surveillance Systems," VS-PETS 2005, Beijing, pp. 15-16.

[3] L.M. Brown, A.W. Senior, Y.L Tian, J. Connell et A. Hampapur, C-F. Shu, H. Merkl et M. Lu "Performance Evaluation of Surveillance Systems under Varying Conditions," PETS 2005, Breckenridge, Colorado, pp.1-8.

[4] F. Bashir et F. Porikli, "Performance Evaluation of Object Detection and Tracking Systems,'' PETS 2006, New-York, pp. 7-14.

[5] EC Funded CAVIAR project/IST 2001 37540, http ://homepages.inf.ed.ac.uk/rbf/CAVIAR/

[6] K. Sohail, I. Umer, S. Saquib et A. Asim, "Bhattacharyya Coefficient in Correlation of Gray-Scale Objects". Journal of Multimedia 1(1): 56-61 (2006)

[7] D. Comaniciu, V. Ramesh and P. Meer, "Real-Time Tracking of Non-Rigid Objects using Mean Shift," in IEEE Proc. CVPR, vol. 2, June, 2000, pp.142-149.